A Publication of the American Psychological Association

Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

ARTICLES

Consistency of Maternal Behavior from Infancy to Preadolescence ... 1
Earl S. Schaefer and Nancy Bayley

Infantile Trauma, Genetic Factors, and Adult Disorganization ... 10
Gardner Lindzey, David T. Lykken, and Harvey D. Winston

Opinion Change as a Function of External Commitment and Amount of Discrepancy from the Opinion of Another ... 16
Milton E. Rosenbaum and Douglas E. Tyler

A Constant Error Approach to the Study of Dimensions of Social Perception ... 25
Leon B. Levy and Robert D. Duran

A Comparison of Personality Structures in American Southern and Northern Culture Groups ... 34
George I. Stein

Patterns of Role Deprivation and Conflict in Response to Parental Separation ... 43
Amerigo Farina

The Validity of Hypnotically Induced Visual Hallucinations ... 50
Harold Wayne Underwood

Personality, Similarity, and Friendship ... 51
Carroll E. Izard

Continued on inside front cover

Volume 61 JULY 1960 Numb

CONTENTS—*Continued*

Control Techniques of Extraversion-Introversion ... 53
Walter J. Perrott

Changes in Perceived Instrumentality in Reputation and Sociometric ... 67
John Sprowls, Oscar A. Parsons, and Ralph Hakstian

Effects of Group Support on the Extinction of an Instrumental ... 73
Edwin H. Silverstein, Seymour B. Sarason, and William H. Smith

Differential Learning and Forgetting as a Function of the Social Frame of ... 82
Robert Thorne
B. Kleinmuntz and L. J. Fox

Interpersonal Knowledge and Reward Leader Potential ... 87
Martin Patchen

Coercive Power and Forces Affecting Conformity ... 94
John R. P. French, Jr., B. William Morrison, and George Levinger

Reinforcement and Conformity under Reward and Punishment ... 102
Morris G. Zelig

Race, Social Class, Religion, and Nationality as Determinants of Social ... 109
Daisy Myers
Harry C. Triandis and Leigh Minturn Triandis

Occupational Choice and Cognitive Functioning: Cultural Restrictions of Ability, ... 117
Interests, and of Ability, Creative Abilities, and ...
J. M. Crooks and C. W. Taylor

Subliminal Perception or Perception of Partial Cues ... 124
Kenton Fraser and Jesse H. Schultz

Cal Cook and Sons ... 129

Ease, Tenseness, and the F Scale ... 135
Morris I. Stein

Mobility in the Mentally Defective ... 141
Fred D. Seiger and James G. Barber

Hostility as Judged by Naïve Interpreters of Projective Responses ... 144
Paul E. Sweet and John Sommer

Taylor Scale and Group Conformity Pressure ... 146
G. J. Mouton, H. Quadermain, and G. Morris Goodman

Personality of Bell Negro-White Intelligence ... 148
Stephen B. Sells

Patterns of Schizophrenic Thought Disorder, Pathogenic Parents, ... 151
M. Perry Feather

Characteristics of Underachieving Gifted Children Identified by Group ... 155
Harold Pilgrim and Carroll Shores

Volume 61 July 1960 Number 1

Journal of Abnormal and Social Psychology

Editor: M. Brewster Smith, *University of California, Berkeley*
Associate Editor: E. J. Shoben, Jr., *Teachers College, Columbia University*

Consulting Editors

Gordon W. Allport
Harvard University

Urie Bronfenbrenner
Cornell University

Roger Brown
Massachusetts Institute of Technology

Donald T. Campbell
Northwestern University

Launor F. Carter
System Development Corporation

Richard Christie
Columbia University

Morton Deutsch
Bell Telephone Laboratories

I. E. Farber
State University of Iowa

Donald W. Fiske
University of Chicago

N. L. Gage
University of Illinois

Roger W. Heyns
University of Michigan

Robert R. Holt
New York University

Irving L. Janis
Yale University

Harold H. Kelley
University of Minnesota

Leonard S. Kogan
Community Service Society of New York

Richard S. Lazarus
University of California, Berkeley

Daniel J. Levinson
Harvard Medical School

David C. McClelland
Harvard University

Leo Postman
University of California, Berkeley

Eliot H. Rodnick
Duke University

Nevitt Sanford
University of California, Berkeley

Seymour Sarason
Yale University

Richard L. Solomon
University of Pennsylvania

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 900 pages. The subscription price per year is $16.00; foreign, $16.50. Single issues are $3.00.

Original manuscripts should be sent to the Editor:

M. BREWSTER SMITH
Department of Psychology
University of California
Berkeley 4, California

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street, N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Mt. Royal and Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Second-class postage paid at Baltimore, Maryland, and at additional mailing offices

Copyright 1960 by the American Psychological Association, Inc.

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 1–6

CONSISTENCY OF MATERNAL BEHAVIOR FROM INFANCY TO PREADOLESCENCE

EARL S. SCHAEFER AND NANCY BAYLEY
National Institute of Mental Health, Bethesda, Maryland

CURRENT social psychological theories of personality development are concerned with the cumulative effect upon the child of the social and emotional behaviors of the parent throughout the period of childhood. Interest in these cumulative effects leads to the question whether from observations or interviews on parental behavior at one developmental stage it is possible to determine consistent, typical parental behavior that existed earlier and will continue throughout childhood. This problem should be of special concern to retrospective studies in which parents of contrasting groups are either observed or interviewed and inferences are then made that current differences in parental behavior existed prior to the children's behavior, or that they contributed to its development.

A second problem in research concerned with evaluating the influence of maternal behavior upon personality development is that of the validity of both observation of mother-child interaction and of the interview as sources of data on the mother's typical behavior with the child. Observation of mother-child interaction is more feasible during infancy when the mother usually maintains close contact with the child and interacts relatively spontaneously with the infant in the presence of the observer. At later ages, when the child's contact with the mother is usually reduced and when the presence of the observer often limits the spontaneity of mother-child interaction, observation is often replaced by an interview with the mother from which inferences are made about her behavior with the child. Little is known about the relationship between observation and interview as sources of data on maternal behavior.

Available longitudinal data permitted tests of two hypotheses that are suggested by these problems:

1. A mother's behavior with her infant is significantly related to her behavior with the same child at preadolescence.

2. Data on a mother's behavior with her child obtained from an interview are significantly related to data obtained from direct observation of mother-child interaction.

The available data led to a research design that unfortunately confounded these two hypotheses. Only two sets of data were available, data quantified from records of observations during the period of one month to three years of the child's life (infancy) and data quantified from interviews with the mother at child's age of from 9 to 14 years (preadolescence).[1] If these two sets of data are significantly related, both hypotheses must be accepted. If, however, the two sets of data are not significantly related, the inconsistency of maternal behavior might be attributed either to true changes in behavior across this time span, to the lack of validity of either the observational or interview data, or both.

METHOD

Subjects and Data

A description of the subjects (*Ss*) and of the research in which these data were collected has been reported elsewhere (Jones & Bayley, 1941). Both infancy observation data and preadolescent interview data were available for mothers of 15 boys and 16 girls. Because of the small size of the two groups the data on mothers of boys and of girls were combined for the analysis of individual scales. The consistency of the two major dimensions were computed both for the combined groups and separately for the two sexes.

The observations were collected by an observer[2] who was present during the testing sessions with mother and child in the University of California Institute of Child Welfare during the first three years of the child's life. Tests with the children were scheduled every month for the first 15 months and every 3 months thereafter through the age of 3 years. The later interviews were made by a skilled psychologist[3] who had no knowledge of the earlier observations. The mothers

[1] The age of the children at the time of the preadolescent interview ranged from 9 to 14 years, with a median and mode of 13 years.

[2] Most of the observations were made by a graduate student in psychology, H. Elizabeth Neall. Some were made by one or other of the examiners.

[3] Pearl Bretnall Meissner.

were interviewed during either one or two visits in the home between the child's ages of nine to fourteen years. The interview was focused upon current family behavior rather than upon earlier events. A comprehensive clinical report of the interviews was written after the home visit. The early observations and later interviews were employed (Schaefer, Bell, & Bayley, 1959) in developing measures for a conceptual scheme that emphasized molar social psychological variables (see Table 1). The conceptual scheme was first used to quantify the birth to 3-year observations and then was modified slightly to make the variables appropriate for the 9- to 14-year interview data. Each of the two sets of data was rated by 3 judges, although only one of these judges rated both sets.

The median combined interjudge reliability (product-moment correlations) for the ratings of the observation data was .82 and the median reliability for the ratings of the interview data was .85. The consistency correlations for the one judge who rated both sets of data were not significantly different from the consistency correlations between independent raters. Only the consistencies of the combined ratings of the three judges will be reported.

In an effort to organize the variables that have been described, the variables that were directly related to the social and emotional behavior of the mother toward her child were intercorrelated for each set of data. Additional variables that were rated but were not included in this analysis are descriptive of the mother's personality, or are related to situational circumstances. An organization of the maternal variables using the circumplex model (Guttman, 1954) and factor analysis (Thurstone, 1947), revealed an ordering within a two-dimensional space (Schaefer, 1959). The two dimensions were labeled Love versus Hostility and Autonomy versus Control for both sets of data.

RESULTS

The infancy observation variables were correlated with the preadolescent interview variables using Pearson produce-moment correlations. Both sets of variables are reported in Table 1 in the circumplex order that was found in the previous analysis (Schaefer, 1959). The combined interrater reliabilities are also reported. All correlations above .36 are significant at the .05 level and all correlations above .47 are significant at the .01 level using a two-tailed test for this sample of 31 *S*s. The significant correlations tend to occur between variables that define a dimension of Love versus Hostility while the insignificant correlations are between variables that define a dimension of Autonomy versus Control at each age level.

As a method of summarizing this finding, the variables that had high intercorrelations and high loadings on a dimension were organized into cluster scores. The Autonomy versus Control dimension was defined for the period of infancy by autonomy, excessive contact, intrusiveness, and fostering dependency and for preadolescence by autonomy, excessive contact and intrusiveness. The Love versus Hostility dimension was defined for the early observations by cooperativeness, positive evaluation of the child, equalitarianism, expression of affection, ignoring, irritability, punishment, punitiveness, perceives the child as a burden, use of fear to control the child, and strictness, and for the later interviews by cooperativeness, positive evaluation of the child, equalitarianism, expression of affection, positive mother-child relationship, ignoring, irritability, punitiveness, use of fear to control the child and rejection of the homemaking role. Sums of the unweighted scores for the variables in a cluster were determined for each of the 31 *S*s at both age levels and these two levels were then intercorrelated, using Pearson product-moment correlations. The intercorrelations and estimated combined interrater reliabilities for each of the clusters are reported in Table 2. The correlation for the combined sexes between the Love versus Hostility clusters is .68. significant beyond the .01 level, while the correlation between the Autonomy versus Control clusters is .26, not significantly different from zero. The results for the two sexes do not differ significantly. The correlation for the Love–Hostility axis for boys is .73 and for girls is .61 and both are statistically significant. The consistency correlation for the Autonomy–Control axis for boys is .18 and for girls is .35 and neither one is statistically significant. These results summarize and confirm the findings reported in Table 1. The correlation of .34 between the Love-Hostility and Autonomy-Control axes at preadolescence merely indicates that the method of clustering variables did not give independent dimensions.

Although Tables 1 and 2 adequately summarize the data on consistency of maternal behavior for variables that clearly relate to the behavior of the mother with the individual child, other variables of the infancy data were analyzed to determine the extent to which they are related to the preadolescent ratings of positiveness of mother-child interaction. The significant correlations of the early ratings

TABLE 1

CORRELATION OF CIRCUMPLEX-ORDERED INFANCY OBSERVATION RATINGS OF MATERNAL BEHAVIOR WITH CIRCUMPLEX-ORDERED PREADOLESCENT INTERVIEW RATINGS OF MATERNAL BEHAVIOR

Preadolescent Interview Ratings	Reliability	Infancy Observation Ratings: Autonomy 80	Positive Evaluation 82	Equalitarianism 84	Expression of Affection 83	Emotional Involvement 87	Fostering Dependency 85	Excessive Contact 67	Achievement Demand 81	Concern About Health 81	Intrusiveness 76	Anxiety of Mother 83	Irritability 82	Punishment 87	Use of Fear to Control 86	Strictness 78	Perceives as Burden 88	Punitiveness 78	Ignoring 89
Autonomy	77	29	01	13	−08	−04	−09	−26	−19	−25	−23	−01	06	−07	−11	−14	08	01	15
Positive Mother-Child Relationship	88	05	*48**	*58*	*39*	25	16	20	03	10	−07	−10	*−51*	*−42*	*−52*	*−36*	*−53*	*−54*	*−41*
Equalitarianism	83	09	*52*	*56*	*45*	28	16	19	09	08	−12	−08	*−53*	*−41*	*−51*	*−39*	*−57*	*−60*	*−43*
Expression of Affection	78	−08	*57*	*63*	*58*	*44*	*39*	29	18	16	10	06	*−51*	*−49*	*−56*	*−44*	*−62*	*−60*	*−56*
Positive Evaluation of Child	84	−07	*60*	*59*	*56*	*40*	*37*	29	12	13	06	−12	*−63*	*−51*	*−58*	*−44*	*−68*	*−62*	*−58*
Emotional Involvement	86	−25	35	*44*	*44*	*40*	*37*	28	*45*	*48*	20	11	−32	*−40*	−33	−29	*−46*	*−44*	*−51*
Achievement Demand	85	−14	22	20	17	14	10	08	31	*46*	12	−09	−29	−26	−20	−09	−30	−21	−28
Fostering Dependency	72	−11	06	16	21	13	22	01	24	24	18	13	−02	−15	−06	−05	−18	−24	−19
Excessive Contact	81	−30	−04	−04	06	01	18	24	10	18	30	06	−04	−05	05	06	−09	−09	−16
Intrusiveness	77	−24	−31	−32	−14	−11	07	04	21	20	34	23	23	08	26	27	18	10	12
Social Isolation of Child	80	−08	−19	*−36*	−22	−26	−11	−04	−02	−10	12	−15	07	17	25	26	07	14	15
Wish to Control Child	88	−18	*−50*	*−49*	*−36*	−24	−06	−06	09	32	33	16	*40*	32	*48*	*42*	*38*	*38*	27
Strictness	77	04	*−52*	*−44*	*−41*	−33	−24	−22	−01	16	12	00	*39*	*38*	*52*	*48*	*44*	*52*	*40*
Use of Fear to Control Child	85	−10	*−67*	*−61*	*−54*	*−36*	−18	−20	−02	07	15	21	*63*	*48*	*64*	*55*	*63*	*66*	*57*
Perceives Child as Burden	81	19	−34	*−37*	*−48*	*−42*	−14	−35	−06	*−46*	−06	−09	01	16	25	27	19	14	26
Irritability	89	−02	*−59*	*−57*	*−47*	−30	−12	−27	06	10	15	29	*64*	*40*	*54*	*38*	*63*	*53*	*51*
Punitiveness	86	−02	*−54*	*−55*	*−45*	−31	−12	−22	−04	01	15	21	*53*	28	*46*	28	*54*	*42*	*42*
Ignoring	89	15	*−63*	*−70*	*−60*	*−49*	*−38*	*−38*	−14	−22	−12	08	*63*	*47*	*60*	*43*	*68*	*64*	*62*

Note.—Decimal points omitted; $N = 31$.

* Significant correlations are italicized; correlations above .36 are significant at .05 level and above .47 are significant at .01 level.

TABLE 2

CORRELATIONS OF CLUSTER SCORES FOR ZERO- TO THREE-YEAR OBSERVATIONS AND NINE- TO FOURTEEN-YEAR INTERVIEWS

(Combined interrater reliabilities in diagonal cells)

Cluster	1	2	3	4
1. Love–Hostility (0–3 years)	(94)			
2. Love–Hostility (9–14 years)	68*	(91)		
3. Autonomy–Control (0–3 years)	−14	−12	(87)	
4. Autonomy–Control (9–14 years)	09	34	26	(84)

Note.—Decimal points omitted; $N = 31$.

* Significant beyond the .001 level of significance using a two-tailed test.

TABLE 3

CORRELATIONS BETWEEN ADDITIONAL ZERO- TO THREE-YEAR MATERNAL RATINGS AND NINE- TO FOURTEEN-YEAR RATINGS OF POSITIVENESS OF MOTHER-CHILD INTERACTION

0- to 3-Year Maternal Behavior Rating	Combined Interrater Reliability	Correlation with 9- to 14-Year Rating of Positiveness of Mother-Child Interaction (Reliability = 88)
Cooperativeness	85	58**
Rejection of the Homemaking Role	83	−45
Poor Physical Health	81	−40
Negative Emotional State	70	−40
Subjective Financial Stress	88	−36
Estimated Intelligence	85	36*
Dominance	84	−24
Mood Swings	79	−24
Self-Abasement	72	19
Overconscientiousness	80	13
Dependency	72	−10
Sociability	52	−04
Narcissism	73	01

Note.—Decimal points omitted; $N = 31$.

* Correlations above .36 significant at .05 level using two-tailed test of significance.

** Correlations above .47 significant at .01 level using two-tailed test of significance.

of cooperativeness of the mother and rejection of the homemaking role with the later rating of positiveness of mother-child interaction, reported in Table 3, can be attributed to the fact that these variables have high loadings on the dimension of Love versus Hostility that is also consistent across this period. Poor physical health of the mother, negative emotional states of the mother, and financial stress are significantly negatively correlated with later positive mother-child interaction. The finding of a positive correlation between estimated intelligence of the mother and later positiveness of mother-child interaction may reflect a halo effect in rating either maternal intelligence or behavior or may reflect a true relationship.

DISCUSSION

The correlation of .68 between the scores of these *S*s for the two sets of data for the dimension of Love versus Hostility, although significantly consistent, may not appear as high as one might wish in order to make inferences about maternal behavior from one age level to another. However, when we consider the many factors that might reduce the correlation between the two sets of data, the size of the correlation is relatively high. There are probably true changes in the behaviors of individual mothers during the 6- to 11-year interval between the observations and the interviews. The two different methods of collecting data, interview and observation, might lead to different impressions of characteristic behavior. The different situations in which the data were collected, the university research setting and the child's home, might influence the behavior of the mothers in different ways. Similar factors may also have contributed to the insignificant correlation between age levels for the Autonomy versus Control dimension. If we assume that the ratings of the Autonomy versus Control dimension are as valid as those for the Love versus Hostility dimension, it would appear that there are greater changes in maternal behavior on this dimension through time. One interpretation of this difference might be that a child's need for love and a positive relationship with the mother is consistent through time, while the child's need for autonomy changes from a need for total dependence and a close relationship at birth to a need for relatively complete autonomy as an adult. Some mothers who have a close relationship

with the child described by the ratings of excessive contact, fostering dependency, low autonomy, and intrusiveness during infancy may shift to give an age-appropriate autonomy as the child matures. Other mothers who have a more distant relationship with the infant may become more possessive, intrusive, and controlling later as the child becomes more independent and more active. Other hypotheses might be that there are individual variations in the degree to which children seek autonomy or there may be changes in the mother's capacity to satisfy the child's needs, and that these changes are reflected in changing scores on this dimension.

The finding that early ratings of poor physical health of the mother, subjective financial stress, and negative emotional states (gloominess, dullness, unresponsiveness) have significant negative correlations with later positiveness of the mother-child-relationship suggests that environmental stresses felt by the mother, reduced physical vitality, or anything which contributes to a negative emotional adjustment of the mother may impair the development of a good relationship with the child. Preoccupation with other problems may reduce the capacity of the mother to provide warmth and security for the child, or these results may be interpreted by the frustration-aggression hypothesis (Dollard, Doob, Miller, Mowrer, & Sears, 1939). Of course, poor health, subjective financial stress, and negative emotional states may all reflect emotional maladjustment of the mother rather than objective stresses.

The findings of this study suggest the need for caution in the interpretation of retrospective studies of the relationship of maternal behavior to the development of children. The relatively low consistency for the dimension of Autonomy versus Control, the control end of which is similar to the concepts of overpossessiveness and overprotectiveness, suggests that it may be difficult to make inferences about early maternal behavior relevant to this dimension from interview data collected at a later date. Of course, the interview data of this study were concerned with current family behavior rather than with attempts to reconstruct what had happened earlier. Although the correlation between the Love versus Hostility scores for the two age levels provided support for making inferences about maternal behavior from one age level to another, the size of the correlation does not preclude, in individual cases, relatively large changes through time.

SUMMARY

An interest in the influence of maternal behavior upon development of the child suggested two hypotheses:

1. A mother's behavior with her infant is significantly related to her behavior with the same child at preadolescence.

2. Data on a mother's behavior with her child collected from an interview are significantly related to data collected from direct observation of mother-child interaction.

Observational data collected during the first three years of the child's life in testing sessions at the University of California Institute of Child Welfare and interview data collected in the home between the child's age of 9 to 14 years permitted a test that confounded the two hypotheses.

The finding of a significant correlation between scores on the two sets of data for a dimension of Love versus Hostility led to acceptance of both hypotheses for that dimension. The finding of an insignificant, although positive, correlation between scores for the dimension of Autonomy versus Control could not be definitely interpreted since this result might be due either to low validity for either the observation or interview data or both, or to true inconsistency in maternal behavior for this dimension. It was found that early ratings of poor physical health of the mother, subjective financial stress, and negative emotional states had significant negative correlations with later ratings of positiveness of mother-child interaction.

REFERENCES

DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., & SEARS, R. R. *Frustration and aggression.* New Haven: Yale Univer. Press, 1939.

GUTTMAN, L. A new approach to factor analysis: The Radex. In P. F. Lazarsfeld (Ed.), *Mathematical*

thinking in the social sciences. Glencoe, Ill.: Free Press, 1954.

JONES, H. E., & BAYLEY, NANCY. The Berkeley growth study. *Child Develpm.*, 1941, **12**, 167–173.

SCHAEFER, E. S., BELL, R. Q., & BAYLEY, NANCY. Development of a maternal behavior research instrument. *J. genet. Psychol.*, 1959, **95**, 83–104.

SCHAEFER, E. S. A circumplex model for maternal behavior. *J. abnorm. soc. Psychol.*, 1959, **59**, 226–235.

THURSTONE, L. L. *Multiple-factor analysis*. Chicago: Univer. Chicago Press, 1947.

(Received July 14, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 7-14

INFANTILE TRAUMA, GENETIC FACTORS, AND ADULT TEMPERAMENT

GARDNER LINDZEY, DAVID T. LYKKEN, AND HARVEY D. WINSTON
University of Minnesota[1]

IN SPITE of a welter of empirical and theoretical activity centering upon the role of infantile experience as a determinant of adult personality it is evident that there are few important issues in this area that have been satisfactorily resolved. Prominent among these open issues is the potential contribution of gene structure to the relation between infantile experience and adult temperament. Although King (1958) has identified gene factors as one of seven major parameters to be considered in studying the relation between infantile experience and adult effects, there is little in the way of compelling empirical findings concerning the role of genetic factors in this setting.

This sparse activity in the world of observation exists in spite of a history of theoretical interest in this issue. There are few general theories of behavior that fail to take into consideration, at least implicitly, the potential impact of constitutional or genetic factors upon early experience and its consequences. Perhaps the most influential and persistent emphasis upon the role of heredity in this context has been provided by Freud (1905) with his discussions of the determinants of fixation, but his position differs from other personality theorists on this issue primarily in regard to explicitness.

The present study was designed to provide further evidence relevant to this general issue. Because of traditional difficulties involved in using human subjects for studies spanning the life history of the organism and the additional requirement that genetic variation be controlled, it was necessary to use subjects drawn from inbred strains of mice.

While the most unusual aspect of this study inquired into the interaction between gene factors and the effects of infantile trauma, it would be a rash investigator indeed who risked his all on an interaction term. Consistent with this premise is the fact that our study also provides information concerning the general effects of infantile trauma, their situational generality and duration, as well as further evidence regarding the influence of genetic factors upon various measures of temperament.

A word should be said concerning previous results that are relevant to the relation between infantile trauma and emotionality, for here, as well as in connection with the interaction effect we have just discussed, we have made specific predictions prior to the study. Clear evidence of an increase in emotionality in mice as a result of infantile trauma has been demonstrated by Hall and Whiteman (1951), while other investigators working with mice (Stanley & Monkman, 1956) and with rats (Griffiths & Stringer, 1952) have failed to find any significant differences in emotionality between control and traumatized animals. One investigator (Ader, 1957) actually has reported results indicating that treated rats were *less* emotional than control animals. Although these studies have led to conflicting results, the firmness of the findings of Hall and Whiteman (1951), coupled with consistent evidence from human investigations (Bowlby, 1951) and a boost from psychoanalytic theory, led us to predict that early traumatic experience would produce an increase in emotionality (timidity).

In summary, this study was designed to provide information concerning:

1. The importance of genetic factors as determinants of emotionality, timidity, and activity.

2. The general effects of infantile trauma, their situational generality, and the duration of these effects through time.

3. The influence of genetic factors upon the relation between infantile trauma and adult temperament.

[1] From the Center for Personality Research. This investigation was supported by a grant from the Ford Foundation. We are deeply grateful to John J. Bittner for his generous assistance and detailed advice concerning the care and breeding of mice. Mary Ahlquist, Samuel Berman, and Arthur Hill all rendered valuable aid at various stages in this study.

METHOD

Subjects

The (*Ss*) of this investigation were offspring of breeding animals belonging to four different inbred strains of mice. The parent animals were obtained from a colony maintained by John J. Bittner of the Cancer Biology Laboratories of the University of Minnesota and represent the outcome of many generations of brother-sister matings. The conventional labels for these strains and the approximate number of generations of controlled brother-sister matings are as follows: C57BL/1 or B/1 (51 generations); C3H (93 generations); DBA/8 or D/8 (27 generations); JK (71 generations). The first three strains were selected on the basis of prior information (Lindzey, 1951) indicating diversity in emotionality, while the JK strain was selected because observation of them in the colony suggested that their behavior was quite different from the other strains. They were also selected so that all four strains could be identified easily by their distinctive coat color.

Within each strain the successive litters at birth were assigned alternately to either the experimental or the control group with the restriction that when the second litter by a particular mother was born that it would be given a different assignment than the first litter. This shifting procedure was continued through all litters until the desired number of animals in either the control or experimental group was achieved, and all subsequent litters were then assigned to the other group until the necessary number was reached. Because of the impossibility of predicting survival rates precisely for a period of almost four months, a variable number of animals resulted in the different experimental groups. The smallest number in any single group was 23 while the largest number was 36. The total number of animals included in the study was 259.

Each litter was raised in a separate cage with the mother until 24 days of age, at which time the young were weaned and segregated by sex. Animals were individually identified by an ear punch code, but this operation was not carried out until completion of the first consequent measure in order to avoid the possibility that this relatively traumatic procedure would wipe out the effects of our experimental treatment.

Infantile Trauma

Control and experimental animals were raised under the same conditions except that at four days of age those animals assigned to the experimental group were on four successive days exposed to an extremely loud, high frequency, auditory stimulus. The procedure followed was essentially the same as that described by Hall and Whiteman (1953) and involved placing the mice on a thin cardboard container in a #1 12-gal. wash tub with a doorbell (#504 Eclipse) fixed to the side. After a 2-min. interval the doorbell was rung for 2 min., following which the mice were left in the tub for another two minutes before being returned to their cages. On the first three trials all animals were placed together in the tub, but on the fourth trial they were separated. The procedure for the control animals was precisely the same except that the bell was not rung while they were in the tub.

Consequent Measures

30-Day Open Field Test. Beginning at 30 days of age each mouse was placed for 2 min. in the same type of wash tub in which the auditory trauma had been administered. The tub was brightly lighted by a shielded 150-w. glazed bulb that was placed directly over the tub at a height of approximately 1½ ft. The mouse was left in the tub for 2 min. at the end of which time the incidence of defecation and urination was recorded and the mouse returned to his living cage. This measure was repeated on 10 successive days. The measure of emotionality was simply presence or absence of defecation and/or urination during each trial.

Stovepipe Test. This instrument was devised by Stone (1929) as a measure of timidity or of the "dominance of hiding tendency over hunger," (p. 36) and was used by Hall and Whiteman (1951) in the study referred to earlier. In this measure the mouse was placed in a starting box connected with a U-shaped stovepipe (each unit approximately 2 ft. in length) that led to a goal box containing a food receptacle holding wet mash. Thus, all that the mouse had to do to reach the food was to enter the stovepipe, follow the passage, and leave the dark pipe for the more brightly illuminated goal box. The mice were again given 10 trials on successive days, commencing at 70 days of age, and they were run under 22½ hrs. of food deprivation. The measure employed was the total time that it took on the 10 trials for the mouse to enter the goal box. The few mice that on particular trials failed to leave the starting box within 5 min. were on these trials arbitrarily assigned the same time as the mouse that took the longest time in reaching the goal box.

100-Day Open Field Test. Commencing at 100 days of age the open field test was repeated for 10 trials with the measure of emotionality consisting of presence or absence of defecation and/or urination summed across the 10 trials for each *S*. In addition a measure of motility was secured by dividing the wash tub into 12 spaces of equal area and counting the number of spaces that each mouse moved through during the two minute trial and again summing the results of the individual trials over the 10 days.

Analysis of Data

The obvious method of statistical analysis for our data was a 2 × 4 analysis of variance and this technique was used for all data except those derived from the first open field test. In the latter case the absence of individual identification made it impossible to summate scores over the 10 trials and, in general, made it virtually impossible to devise a sensitive measure of the interaction hypothesis. Demonstration of strain differences and treatment effects was not particularly difficult, in part because they were relatively evident, and here we have relied upon descriptive presentation and the use of such simple devices as the sign test. The best estimate of interaction that could be devised, however, was to adapt the chi square test and then to apply this test separately to the data for each of the 10 days. This is obviously a somewhat unsatisfactory test of the interaction. We have not hesitated to apply one-tailed

tests of significance to data related to directional hypotheses stated in advance of the study.

For purposes of analysis of variance we reduced all groups to the size of the smallest group in order to avoid the complexities of disproportionality. *Ss* were eliminated on a chronological basis beginning with the animals that were last to complete the procedures. In the case of descriptive presentation of results we have included all animals studied with an appropriate indication of the number in each group.

RESULTS AND DISCUSSION

Gene Determinants of Temperament

Here we are concerned with strain differences in two open field measures of emotionality, the stovepipe measure of timidity, and an open field measure of motility. The findings are clear and dramatic. No sentient person could observe the data reported in Tables 1–6 and conclude for any one of these attributes that the animals in all the strains could have been drawn from a single common population. In some instances the strains are so different as to show virtually no overlap. In view of the fact that the only known difference between these strains lies in their genetic constitution, it seems a prudent conclusion that genetic factors make a considerable contribution to each of our three measures of temperament.

Our findings are given added weight when coupled with the results of a number of earlier investigations using both rats (Hall, 1938; Broadhurst, 1958) and mice (Lindzey, 1951; Thompson, 1953, 1956; McClearn, 1959), all of which demonstrate wide strain differences on comparable measures. More interesting than this general congruence is a specific examination of the incidence of defecation on the part of the three strains included in both the present study and the previous investigation by Lindzey (1951). In each study, C57, DBA, and C3H strains were employed and in each case an open field test of emotionality was utilized. In Table 2 the percentage of animals defecating in each strain in the earlier study, and the percentage of defecation over 10 trials in each strain in the present study, are compared. The figures are highly constant in spite of the fact that the data were collected with more than a decade of time intervening, with considerable variation in the details of the open field test, with different persons collecting the data, and with animals that are not even known to belong to the same subline within the various strains. Even worse, we *do* know that the two samples were drawn from colonies that have been separated long enough to provide ample opportunity for "genetic drift" as well as spontaneous mutation. Given all of these sources of potential variation it is quite surprising to find the consistency of the strains and the measure so great. Thus, we are afforded the luxury of simultaneously gaining confidence in the stability of the organism we are studying and the reliability of our measure.

In general, these findings serve to emphasize again the pervasive impact upon behavior of genetic factors and thus strongly suggest the inadequacy of attempts to formulate or ac-

TABLE 1
PERCENTAGE OF ANIMALS DEFECATING DAILY ON 30-DAY OPEN FIELD TEST

Strain	Treatment	*N*	Day										Average
			1	2	3	4	5	6	7	8	9	10	
C57BL/1	Control	28	28.6	28.6	32.1	53.6	35.7	39.3	46.4	32.1	28.6	17.9	34.3
C57BL/1	Experimental	25	32.0	36.0	40.0	40.0	60.0	52.0	36.0	40.0	36.0	32.0	40.4
C3H	Control	34	85.3	88.2	82.3	64.7	73.5	70.6	85.3	73.5	73.5	73.5	77.1
C3H	Experimental	39	92.3	92.3	92.3	79.5	82.0	84.6	92.3	94.9	94.9	100.0	90.5
DBA/8	Control	37	64.9	81.1	73.0	70.3	62.2	64.9	62.2	64.9	59.5	67.6	67.0
DBA/8	Experimental	35	54.3	80.0	82.9	80.0	82.9	82.9	74.3	57.1	77.1	71.4	74.3
JK	Control	29	58.6	79.3	82.8	86.2	100.0	89.6	96.5	72.4	93.1	86.2	84.5
JK	Experimental	32	90.6	87.5	84.4	90.6	93.8	96.9	90.6	87.5	87.5	93.8	90.3

TABLE 2
CONSISTENCY OF EMOTIONALITY MEASURE IN 1951 STUDY AND PRESENT STUDY

	C57BL	C3H	DBA
Percent defecation (1951)	37 (N = 100)	72 (N = 100)	68 (N = 100)
Percent defecation (present study)	34 (N = 28)	77 (N = 34)	67 (N = 37)

TABLE 3
ANALYSIS OF VARIANCE OF STOVEPIPE TEST

Source	df	Sum of Squares	Mean Square	F	P
Treatment	1	2,350,782	2,350,782	1.93	<.10*
Strain	3	241,013,412	80,337,804	65.96	<.0001
Interaction	3	10,241,048	3,413,682	2.80	<.05
Within (error)	176	214,356,579	1,217,935		
Total	183	467,961,821			

* One-tailed test.

count for behavior that do not direct careful attention to the role of hereditary factors. It is interesting to note that it is not necessary to inbreed selectively for particular characteristics in order to produce clear evidence for the influence of genic factors upon behavior. Inbreeding of the present animals was uninfluenced by the temperament characteristics of interest in the present study, and, in spite of this, we find the particular genetic patterns that have been isolated are dramatically different in their behavioral consequences.

Influence of Infantile Trauma upon Temperament

What exactly were the effects of the auditory, infantile trauma? The reader will recall our prediction that this experience would lead to an increase of emotionality and timidity. Examination of the data summarized in Table 1 reveals that in spite of the awkwardness resulting from our inability to identify individual animals, there is little doubt that the experimental treatment led to an increase in emotionality as measured by the 30-day test. Inspection of the proportion of experimental and control animals defecating on each of the 10 trials within the four strains reveals that 32 of the 40 comparisons point to the greater emotionality of the traumatized animals. This incidence of "hits" is well beyond the 1% level of significance, even utilizing the relatively insensitive sign test. Actually, the within-strain results for the C3H animals (on all 10 trials the experimental animals showed a greater proportional defecation) are by themselves significant at below the 5% level using the sign test, while results for the other strains (7 or 8 trials out of 10 in which experimentals defecate more) approach significance.

The results of the stovepipe test, summarized in Table 3, provide suggestive evidence ($p < .10$) for a treatment effect indicating greater timidity (slower in getting to reward) on the part of the traumatized mice. Consistently, we find in Table 4 that at the 100-day open field test there is also a difference in emotionality indicating that the treated animales were more emotional. Only in the case of motility, as revealed in Table 5, do we find little indication of treatment effects upon behavior.

To summarize, there seems little doubt that subjecting infant mice to a noxious, auditory stimulus of high intensity leads to an increase in emotionality in adulthood and there is some basis for believing that there is also an increase in timidity. There is no evidence, however,

TABLE 4
ANALYSIS OF VARIANCE OF 100-DAY OPEN FIELD TEST: EMOTIONALITY

Source	df	Sum of Squares	Mean Square	F	P
Treatment	1	16.5093	16.5093	2.86629	<.05*
Strain	3	679.4458	226.4819	39.3211	<.0001
Interaction	3	2.1082	.7027	—	—
Within (error)	176	1013.7287	5.7598		
Total	183	1711.7920			

* One-tailed test.

TABLE 5
ANALYSIS OF VARIANCE OF 100-DAY OPEN FIELD TEST: MOTILITY

Source	df	Sum of Squares	Mean Square	F	P
Treatment	1	14,635.2	14,635.2	1.29	<.50
Strain	3	1,421,619.4	473,873.1	41.8	<.0001
Interaction	3	24,268.4	8,089.5	.071	—
Within (error)	176	1,994,390.6	11,331.8		
Total	183	3,434,213.6			

TABLE 6
MEANS AND *SD*s FOR STOVEPIPE TEST AND 100-DAY OPEN FIELD TEST

Strain	Treatment	Stovepipe		100-Day Emotionality		100-Day Motility	
		Mean (N = 24)	*SD*	Mean (N = 23)	*SD*	Mean (N = 23)	*SD*
C57BL/1	Control	2301	1339	4.26	3.40	401.2	135.3
C57BL/1	Experimental	2069	882	4.57	3.19	355.0	165.1
C3H	Control	3322	1669	8.43	1.58	198.2	110.7
C3H	Experimental	4312	1039	9.35	1.05	188.1	85.9
DBA/8	Control	773	436	6.39	2.84	272.5	63.3
DBA/8	Experimental	778	407	6.96	2.71	287.7	99.7
JK	Control	1480	1061	8.87	1.51	163.3	67.1
JK	Experimental	1601	1007	9.48	1.02	133.0	53.7

that this experience influences motility in an open field setting.

These findings confirm our prior expectations and fit very neatly with the results reported by Hall and Whiteman (1951) on the basis of similar experimental operations. However, our results directly contradict the conclusion of Ader (1957), based upon an investigation dealing with rats, and are inconsistent with the findings of Griffiths and Stringer (1952), and Hunt and Otis (1955) based on observation of rats, as well as the findings of Stanley and Monkman (1956) derived from the study of mice. There are so many parametric differences between all of these studies (species of animal, nature of trauma, age at trauma, age at consequent measure) that it is probably useless to attempt a rational analysis of the basis for these experimental differences. It is worth note, however, that there is actually no evidence in any study for infantile trauma *decreasing* emotionality. Although Ader has suggested this possibility, his own data, in the only study in which he employed stimuli that could reasonably be considered noxious or traumatic (Ader, 1959), suggest a heightening of emotionality as a result of infantile trauma. Moreover, Ader's attempts to account for the findings of Hall and Whiteman on the basis of the fact that they used seizure-susceptible mice and an infantile trauma identical to that used in producing audiogenic seizures, is not supportable in view of the fact that the present study led to comparable results with strains such as the C57BL/1 and C3H, which are very low in seizure incidence (Lindzey, 1951).

Duration of traumatic effects. Because of the imprecision of our measures and the differences in the form of our data at age 30 days and age 100 days, it is impossible to make any exact estimate of the extent to which the experimental effects may have diminished with the passage of time. However, it is evident that at 30 days, at 70 days, and at 100 days, the influence of the infantile trauma is clearly observable. Whatever may have been lost in intensity, the main effect is still manifest at the time of the last measure. Actually, a close descriptive inspection of the results of the present study, and the results contained in the Hall and Whiteman paper, suggest that there may have been some diminishing of treatment effects by the time of the 100-day test, but there is no firm evidence to support this contention.

Generality of traumatic effects. One obvious and important question has to do with the narrowness with which the effects of the infantile trauma are linked to situations resembling the original traumatic setting. Obviously, the implications of the trauma for the organism are very different if it leads to a generalized response disposition that is elicited in many settings, rather than leading to effects that can be observed only in situations very similar to the original one. Both Beach and Jaynes (1954) and Ader (1959) attempt to account for the increased emotionality following infantile trauma reported by Hall and Whiteman on the basis of the similarity between the traumatic setting and the setting in which emotionality was measured. The results of the

present study are quite specific in refuting this proposal, as we secured consistent empirical findings in connection with both Stone's measure of timidity and Hall's open field measure. The stimulus differences between the stovepipe test and the open field test are so profound that if we are willing to concede the possibility of stimulus generalization from the original setting to the stovepipe test, we have already granted tremendous generality to the effects of the trauma. In brief, then, our findings suggest that the results of the infantile trauma are of a relatively general nature and should be discernible in a wide variety of different settings.

Impact of Genetic Variation upon Effects of Infantile Trauma

We have already seen that our behavioral measures of temperament are heavily influenced by genetic variation and that these same measures covary with traumatic, infantile experience. The essential question remaining is whether these two sets of determinants show any interaction. Are the changes in temperament that can be attributed to early infantile experience in part dependent upon the gene structure of the organism undergoing the experience?

It was impossible to devise a completely satisfactory test of the interaction effect for the 30-day open field test of emotionality because the individual animals were not identified. Consequently, the best we could do was to measure the interaction between strain and treatment effects for each of the 10 trials individually. The resulting χ^2 coefficients provide no evidence to support the hypothesis of interaction. It is true that we would have preferred to use a single more sensitive measure of emotionality (incidence of defecation summed over 10 days), and it is possible to speculate that with this increased sensitivity we would have been able to detect an interaction. There is, however, no direct evidence to support this contention.

Our second measure of interaction between strain and treatment effects deals with the stovepipe measure of timidity and, as the results summarized in Table 3 indicate, we found clear evidence for the existence of an interaction. These results suggest that changes in timidity resulting from the infantile trauma are in part dependent upon the strain (genetic makeup) of the mouse. In particular, the C57BL/1 mice seemed much less influenced by the experimental treatment than the other strains. The third test of the interaction hypothesis is presented in Tables 4 and 5, where we find no evidence for an interaction between strain and treatment in effects upon emotionality or motility at the 100-day test.

The simplest conclusion that can be derived from these findings is that there is an interaction between gene factors and infantile trauma for timidity, as measured by the stove pipe test, but not for emotionality or motility. While this finding may ultimately be substantiated, there are certain unsatisfactory aspects to our test of the interaction of strain and infantile trauma in the case of the two emotionality measures and the measure of motility. The test for the 30-day data had to be carried out upon single observations rather than observations cumulated over 10 days and the test of interaction for the 100-day data may have suffered from a diminishing of treatment effects at this stage. It is known that the analysis of variance measure of interaction is not a powerful test and it is possible that interaction could be demonstrated only with more powerful treatment effects or a larger number of *S*s. Our failure to find treatment effects upon motility made our interaction measure in this area relatively meaningless.

So far as our hypothesis is concerned, we have obtained evidence for the existence of interaction between strain and infantile treatment effects as predicted, but this relationship was not so general as we had expected. Our conviction concerning the existence of some degree of interaction is strengthened by the results obtained by King (1959), demonstrating an interaction between the effects of infantile handling and membership in different subspecies, and Valenstein, Riss, and Young (1955), suggesting a relationship between the effects of isolation upon sexual behavior and membership in a particular inbred strain. An additional study by King (1957) also provides evidence suggestive of such an interaction, although the experimental treatment (isolation) did not take place in this case until after weaning. The findings indicate that changes in aggressive behavior that were produced by isolation in one mouse strain were not dupli-

cated in a second strain. There is unfortunately, no direct test of the difference between the two strains in treatment effects. Joint consideration of all of these findings provides a relatively firm basis for concluding that the consequences of infantile trauma are in part dependent upon the genetic structure of the organism experiencing the trauma.

What are the implications of such a finding for psychological theory? One may contend that these results, even secured in connection with mere mice or guinea pigs, provide a type of confirmation of Freud's assertion that any attempt to map early experience into adult behavior must allow for the contribution of genetic factors. While such a statement has a rather hollow sound at the level of human behavior because of the little that is known concerning the gene structure of man, it does have some specific empirical implications. For example, our results would suggest that when an investigator works with *S*s of unknown or uncontrolled heredity, it is altogether possible to conduct an otherwise exemplary study of the effects of infantile experience and fail to find evidence for such effects. Or, more generally, we may expect that a variety of different empirical findings might be observed in similar studies as a consequence of investigators dealing with *S*s of various genetic backgrounds, rather than as a result of faulty experimental technique. In general, this is a finding that makes life more complex for both investigator and theorist, and in an area where there has never been any shortage of complexity.

SUMMARY

This investigation was concerned with the effects of infantile trauma upon adult temperament, the influence of genetic factors upon temperament, and the possibility of an interaction between early experience and gene structure in their influence upon adult behavior.

Four strains of homozygous mice were used in the study, with the infant offspring assigned by litter to either an experimental or control group. In all, 259 mice were studied. The experimental mice, beginning at four days of age, were exposed on four successive days to a noxious, auditory stimulus while the control animals were treated in an identical manner except that they were not exposed to the traumatic stimulus. At 30 days of age all mice were examined for 10 successive days in an open field test of emotionality and beginning at 70 days of age the mice were again observed for 10 successive days in a stovepipe test of timidity. A final measure of emotionality, and a measure of motility, were secured from 10 days of observation in an open field test beginning at 100 days of age.

The data obtained provide clear and compelling evidence for the importance of genetic factors as determinants of emotionality, timidity, and motility. The four strains displayed marked differences in all three attributes. There was also direct evidence for the influence of infantile trauma upon emotionality and suggestive evidence in regard to timidity, but no evidence for such an influence upon motility. The effects of the infantile trauma were enduring, extending at least to an age of 100 days, and were not limited to stimulus situations closely similar to the orginal traumatic situation. We found evidence of an interaction between the effects of infantile trauma upon stovepipe timidity and genetic factors, but there was no evidence for such an interaction in the case of emotionality. All of our positive findings are supported by evidence supplied by other investigators.

These findings not only demonstrate the central developmental importance of genetic factors and infantile trauma, they also underline the relative complexity of the relationship between infantile experience and adult behavior. Given an interaction between gene factors and infantile trauma it is readily understandable that inconsistent results might be observed by investigators working with heterozygous *S*s of unknown gene structure. Finally, we have pointed to the consistency between the present findings and Freud's formulations concerning the role of constitutional factors in the developmental process.

REFERENCES

ADER, R. Effects of early experience on emotionality. *Amer. Psychologist*, 1957, **12**, 410.

ADER, R. The effects of early experience on subsequent emotionality and resistance to stress. *Psychol. Monogr.*, 1959, **73**(2, Whole No. 472).

BEACH, F. A., & JAYNES, J. Effects of early experience upon the behavior of animals. *Psychol. Bull.*, 1954, **51**, 239–264.

BOWLBY, J. *Maternal care and mental health*. Geneva: World Health Organization, 1951.

BROADHURST, P. L. Determinants of emotionality in the rat: III. Strain differences. *J. comp. physiol. Psychol.*, 1958, **51**, 55–59.

FREUD, S. Three essays on the theory of sexuality. (Originally published 1905.) In *The standard edition of the complete psychological works of Sigmund Freud*. Vol. 7. London: Hogarth, 1953, Pp. 125–248.

GRIFFITHS, W. J. JR., & STRINGER, W. F. The effects of intense stimulation experienced during infancy on adult behavior in the rat. *J. comp. physiol. Psychol.*, 1952, **45**, 301–306.

HALL, C. S. The inheritance of emotionality. *Sigma Xi Quart.*, 1938, **26**, 17–27.

HALL, C. S., & WHITEMAN, P. H. The effects of infantile stimulation upon later emotional stability in the mouse. *J. comp. physiol. Psychol.*, 1951, **44**, 61–66.

HUNT, H. F., & OTIS, L. S. Restricted experience and "timidity" in the rat. *Amer. Psychologist*, 1955, **19**, 432.

KING, J. A. Relationships between early social experience and adult aggressive behavior in inbred mice. *J. genet. Psychol.*, 1957, **90**, 151–166.

KING, J. A. Parameters relevant to determining the effect of early experience upon the adult behavior of animals. *Psychol. Bull.*, 1958, **55**, 46–58.

KING, J. A., & ELEFTHERIOU, B. E. Effects of early handling upon adult behavior in two subspecies of deermice, *Peromyscus maniculates*. *J. comp. physiol. Psychol.*, 1959, **52**, 82–88.

LINDZEY, G. Emotionality and audiogenic seizure susceptibility in five inbred strains of mice. *J. comp. physiol. Psychol.*, 1951, **44**, 389–393.

MCCLEARN, G. E. The genetics of mouse behavior in novel situations. *J. comp. physiol. Psychol.*, 1959, **52**, 62–67.

STANLEY, W. C., & MONKMAN, J. A. A test for specific and general behavioral effects of infantile stimulation with shock in the mouse. *J. abnorm. soc. Psychol.*, 1956, **53**, 19–22.

STONE, C. P. Wildness and savageness in rats of different strains. In K. S. Lashley (Ed.), *Studies in the dynamics of behavior*. Chicago: Univer. Chicago Press, 1932. Pp. 3–55.

THOMPSON, W. R. The inheritance of behavior: Behavior differences in fifteen mouse strains. *Canad. J. Psychol.*, 1953, **7**, 145–155.

THOMPSON, W. R. The inheritance of behavior: Activity differences in five inbred mouse strains. *J. Hered.*, 1956, **47**, 147–148.

VALENSTEIN, E. S., RISS, W., & YOUNG, W. C. Experiential and genetic factors in the organization of sexual behavior in male guinea pigs. *J. comp. physiol. Psychol.*, 1955, **48**, 397–403.

(Received May 25, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 18-20

OPINION CHANGE AS A FUNCTION OF EXTERNAL COMMITMENT AND AMOUNT OF DISCREPANCY FROM THE OPINION OF ANOTHER[1]

MILTON E. ROSENBAUM AND DOUGLAS E. FRANC
State University of Iowa

A CONSIDERABLE amount of research evidence (Hovland, Janis, & Kelley, 1953; Osgood & Tannenbaum, 1955) has indicated that an attempt to influence change of the opinions of an audience is facilitated by a positive evaluation of the communicator. One strategy that a communicator may employ to gain a favorable appraisal is to present his conception of the audience's current position on a relevant dimension of opinion. This technique is frequently seen as setting the stage for acceptance of the communicator as a discerning and straightforward person (Yeager, 1951) or as a means of establishing a tie to the audience (Powell, 1951). On some occasions, such attribution of position may interfere with the communicator's intention rather than furthering it. The purpose of the present study is to investigate some of the conditions under which attribution of position affects the response to an attempt to change opinions.

In an earlier study (Rosenbaum & Zimmerman, 1959) dealing with an attempt to induce change in the opinions of a sample of students in a Southern university concerning segregation, it was found that attribution of a position that was congruent with the previously recorded opinions of the *S*s resulted in significantly greater resistance to a persuasive communication in comparison with the change of opinion found among control *S*s to whom no attribution was made. No significant effect was found for an incongruent attribution. The postitive finding was interpreted as involving a process tentatively labeled *external commitment*. The term "external commitment" is chosen on analogy to the conventional variable of *self-commitment*. Self-commitment is typically evoked by having the person indicate publicly an opinion or judgment prior to exposure to some procedure intended to influence his views. Results of several studies (Deutsch & Gerard, 1955; Fisher, Rubinstein, & Freeman, 1956; Hovland, Campbell, & Brock, 1957) indicate that resistance to the succeeding influence attempt accompanies this initial verbalization of position.

External commitment can be hypothesized to function similarly to self-commitment. It too should produce resistance to influence attempts when the attribution is *congruent* with the person's current response tendency and opposite to the direction of the influence attempt. External commitment procedures also allow attribution of a position that is *incongruent* with an individual's current response and in this case, susceptibility to an attempt to change opinion should be facilitated. In the present study, conditions are arranged by means of which congruent and incongruent external commitment procedures can be compared with a control or no commitment condition.

Evaluation of the earlier study (Rosenbaum & Zimmerman, 1959) led to the hypothesis that the failure to demonstrate the facilitative effect of incongruent commitment may have been due to the deep-seated and emotional character of the opinion area studied or to the extreme amount of change advocated or both. In the present study, varying amounts of discrepancy between *S*'s opinion and the opinion of the influencing agent are employed in order to deal directly with the second of the suggested factors. Apparently less deep-seated opinions are studied but no attempt is made to manipulate this variable directly.

The relationship between amount of opinion discrepancy and amount of change obtained has been studied by Goldberg (1954) and Hovland and Pritzker (1957) with the results of their research indicating a positive relationship; the greater the opinion discrepancy, the

[1] The authors wish to express their gratitude to Harold P. Bechtoldt and Dee W. Norton for statistical and editorial assistance.

greater the amount of change obtained.[2] A portion of the present study can be considered a replication of Hovland and Pritzker's since similar items and influence procedures are employed. Hovland and Pritzker recognize with appropriate caution[3] that the positive relationship obtained may apply only to certain types of opinion items and up to certain degrees of opinion discrepancy. Because of the possibility that the inhibiting or facilitative effect of external commitment is contingent on the amount of opinion discrepancy in the persuasive attempt, a factorial design is used by means of which potential interactive effects can be evaluated.

METHOD

Approximately 400 students in an introductory psychology class responded to a questionnaire that contained, among other items, adaptations of the 12 opinion items employed by Hovland and Pritzker (1957). A sample item reads "All things considered, Washington was a greater president than Lincoln." The questionnaire required response on a six-point scale ranging from "strongly agree," indicated by a +3, to "strongly disagree" which was represented by a −3. From this larger population, 72 *S*s "volunteered" to participate in the experimental session as part of their required hours of extra-class experimental participation.

The experiment was presented to the *S*s in individual sessions as being concerned with social perception and they were instructed to write a one-page essay on the topic "The Value of a College Education." Upon completion of the essay, *E* took it from the room informing *S* that the essay was to be read by a professional person, who would, on the basis of his reading, make judgments as to the *S*'s opinion on other topics. *S* was told that the professional people who were participating in the experiment "are highly competent in their own fields and have wide knowledge in other fields" and that the major focus of the study was to evaluate the social perceptiveness of people who have to deal with others.

A brief delay ensued during which *E* was ostensibly dealing with the professional person. *E* then returned with a booklet containing 24 pages. Each page had at the top the statement of an issue and below that three rows of six boxes. Each box represented a point on the agree-disagree continuum. The *S*s were specifically informed on a face sheet and via oral instruction that the first line of boxes on each sheet represented the professional person's judgment of *S*'s opinion on the particular issue at the top of the sheet. The second line of boxes on each sheet represented the professional person's own position and the third line was to be used by the *S* to indicate his own current opinion. Twelve pages of the booklet contained the critical items that had appeared in the earlier questionnaire and 12 pages presented filler items. A standard order of presentation was employed that permitted no more than two critical items to appear together.

Employing this form, the top line was used in order to administer the external commitment variable. The second line served for the opinion discrepancy variable and the third line represented the dependent responses to be evaluated along with the prequestionnaire responses.

The commitment variable was carried out in the following manner. One-third of the *S*s were placed in a condition of congruent commitment in which all 12 critical items were checked on the first line at the position these *S*s had indicated in their prequestionnaires. One-third of the *S*s were assigned to a condition of incongruent commitment in which check marks were placed three boxes away from the *S*'s prequestionnaire position thereby clearly attributing to him a position opposed to his original position. For example, if *S* had intially strongly disagreed with an item, he was attributed to a "slightly agree" position. The remaining one-third of the *S*s were exposed to a control condition in which no check marks appeared on the first line. These *S*s were told that the professional person's predictions were made on separate sheets and the presence of the top line on their sheets was to facilitate later scoring procedures.

The procedures to vary opinion discrepancy were carried out on the second line. One-third of the *S*s in each of the commitment conditions were exposed to a check mark one box away from *S*'s original position, one-third to a two-unit difference and the remaining third to three units. In all cases, the check mark on the second line was placed in a direction toward the opposite side of the agree-disagree dichotomy from that indicated by *S*'s original position. The 12 filler items were checked in a standard manner that precluded the repetition of similar check patterns throughout the booklet.

The above procedures result in a 3 × 3 factorial design with eight *S*s per cell. *S*s were assigned randomly to the resulting nine combinations of procedures manipulating commitment and opinion discrepancy.

The dependent variable, amount of change, was

[2] The use of the term "advocated change" in Hovland and Pritzker's (1957) study and others using similar operations is not completely consistent with the operations employed. In these studies, as in the present investigation, the *S*s are presented with the judgment or opinion of another person which is to varying degrees different from the *S*'s judgment. In employing this operation, no additional statements are made directing, recommending, suggesting or in any other way indicating the desirability of change on the part of the *S*s. Although advocacy of change may be implicit in the presentation of the other person's response, only empirical evidence can demonstrate that this is so. A forthcoming paper will consider the possibility that different processes are involved in change produced by opinion discrepancy and by advocacy of change.

[3] Fisher and Lubin (1958) present data indicating that with extreme amounts of opinion discrepancy little change is obtained.

derived by relating the *S*'s response on the third line of the critical booklet items to his prequestionnaire responses to the same item. The six-point Likert scale was given unit values ranging from a score of one for "strongly agree" to a score of six for "strongly disagree". A difference score for an item was obtained by subtraction of the final score from the prequestionnaire score. The sign of the change was computed by assigning a plus to changes in the direction of the professional person's own opinion and a minus for changes away from the professional person's position.

RESULTS

After Bartlett's test indicated no basis for rejecting the assumption of homogeneity of variances, an analysis of variance was performed on the change scores. Table 1 summarizes this analysis and reveals that both main effects and the interaction are significant at acceptable levels.

The significant mean square for interaction suggests that the relative effects of opinion discrepancy are in some way dependent on external commitment. Statistical tests indicate no reliable differential effects of opinion discrepancy under conditions of either congruent or incongruent commitment, although the data in Table 2 suggest a slight increase in actual change with increasing magnitude of opinion discrepancy. Under the control condition, however, the mean change (1.56) for for the 3-unit opinion discrepancy condition is significantly larger ($p < .01$) than the mean changes (.63 and .73) for the 1-unit and 2-unit opinion discrepancy conditions.

In spite of the reliable interaction effect, the magnitude of the F ratio (27.94) for commitment conditions justifies the inference that this factor produces reliable differences in mean change under all three degrees of opinion discrepancy. The average change (.97) under the control condition is significantly different from the average change (.54) under the congruent condition and from the average change (1.49) under the incongruent condition. The direction of these differences supports the contention that congruent external commitment produces, in general, resistance to change and that incongruent commitment tends to facilitate change.

Inspection of overall mean changes under the three conditions of opinion discrepancy reveals the same pattern found by Hovland and Pritzker (1957), i.e., the greater the opinion discrepancy the greater the actual change.

TABLE 1
ANALYSIS OF VARIANCE OF OBTAINED CHANGE SCORES

Source	*df*	Mean Square	*F*	*P*
External commitment	2	784.04	27.94	<.001
Opinion discrepancy	2	216.17	7.70	<.01
Interaction	4	71.46	2.55	<.05
Within	63	28.06		

TABLE 2
MEAN OBTAINED CHANGE

External Commitment	Opinion Discrepancy Units			
	1	2	3	Combined
Congruent	.45	.54	.63	.54
Control	.63	.73	1.56	.97
Incongruent	1.26	1.58	1.63	1.49
Combined	.77	.95	1.27	

As previously noted, however, this pattern is associated primarily with the commitment control condition and even there it is reliably established only between the conditions of 2-unit and 3-unit discrepancy. This last result was obtained with a considerably smaller *N* than Hovland and Pritzker employed.

Evaluation of amount of change obtained in terms of the initial positions held by the *S*s yielded essentially the same pattern of results for those holding extreme, moderate, or more neutral initial positions. Similar results were also obtained when differences in directionality of initial opinions were examined.[4] Previous research (Hovland, Janis, & Kelley, 1953) has indicated that resistance to change is frequently manifested by contracommunication or "boomerang" responses. Table 3 presents the frequency with which individuals made one or more boomerang responses in each combination of conditions. Chi square analysis indicates a nonsignificant effect of the various conditions employed on boomerang responses.

Perhaps, it might be suggested, the final responses of the *S*s can be interpreted without

[4] Statistical analyses of these results are precluded because of extreme disproportionality of cell entries and absence of a control group to account for possible regression.

TABLE 3
FREQUENCY OF APPEARANCE OF ONE OR MORE BOOMERANG RESPONSES PER *S*

External Commitment	Opinion Discrepancy Units		
	1	2	3
Congruent	7	7	4
Control	4	6	4
Incongruent	5	3	3

TABLE 4
ANALYSIS OF VARIANCE OF DEVIATION SCORES

Source	*df*	Mean Square	*F*	*P*
Congruence	1	105.13	10.67	<.01
Separation of check marks	1	94.53	9.60	<.01
Interaction	1	4.50	—	—
Within	28	9.85		

reference to initial positions; that is, if the final responses are related to attribution and opinion discrepancy, statements concerning external commitment may be gratuitous. Certain conditions within the present experiment offer data that make possible an examination of this suggestion. Under the condition of congruent commitment and one-unit discrepancy and that of incongruent commitment and two-unit discrepancy, *S*s were presented with patterns of check marks that were essentially identical. The same similarity obtained between the conditions of congruent commitment and two-unit discrepancy and of incongruent commitment and one-unit discrepancy. Along with the variation of congruence these two sets of conditions differed in the amount of separation of the check marks representing attribution and opinion discrepancy. In the former conditions one unit separated the two check marks, while in the latter conditions a two-unit separation was presented.

If final response is related only to the pattern of check marks no difference should appear between conditions that are differentiated only by the labels, congruent and incongruent. If differences between congruent and incongruent commitment conditions do appear, reference to initial positions is appropriate.

In order to test these alternatives the four conditions indicated were analyzed in a 2×2 factorial design in which one independent variable can be described in terms of congruence and the other in terms of separation of check marks. The dependent measure was the average absolute deviation for each *S* of final responses from the midpoint between the two check marks representing the attribution and the opinion discrepancy. Table 4 presents the summary of the analysis of variance of the deviation scores.

The results indicate that congruence is related to final responses. Within the limits of the present data, it appears that initial position is appropriately considered in evaluation of final responses and that statements concerning external commitment are not gratuitous. Final response is also related to the amount of separation between the two check marks, which is in part reflective of the effect of opinion discrepancy.

DISCUSSION

The results of the present experiment taken together with those of the earlier study (Rosenbaum & Zimmerman, 1959) support the contention that the attribution of an opinion that is congruent with a person's current response tendency produces resistance to a subsequent attempt to influence his opinion. These results are comparable to those obtained under self-commitment procedures, and therefore the process would seem to be appropriately named *external* commitment.

In accounting for the positive results of the earlier study, use was made of the interpretation employed by Hovland et al. (1957) as to the basis for commitment effects. Consistency of behavior is viewed as a desirable characteristic for which socially mediated rewards are dispensed. Under self-commitment procedures, the person himself states the standard to which he is to remain consistent. Failure to do so is frequently socially punished by such terms as "inconsistency," "wishy-washy," "mixed-up," "neurotic," etc. In situations involving external commitment the standard is set by the source of the attribution and can be regarded as a cue to which a matched response has in the past led to social reward for most persons in our culture. In effect, the external source is stating a standard

for a particular response on the part of the recipient of the attribution. However, in the previous study (Rosenbaum & Zimmerman, 1959) and in the current experiment, conflict arises from the fact that not only does the external source present a standard in the form of the attribution, but he provides a second standard in presenting his own position. Potential approval would seem to be offered for presenting a response similar to that of the external source as well as for responding in accord with the attribution. Whether the attribution standard or the discrepancy standard governs the final response presumably depends on undetermined factors associated with the person's response hierarchy. For an unselected group the mean change should appear as a compromise response falling between the two standards. An increase in obtained mean change should then occur, varying directly with the amount of opinion discrepancy. Although a tendency toward a positive relationship is indicated by the results for different amount of opinion discrepancy in the congruent commitment conditions, the relationship is clearly not significant, in spite of evidence in the no commitment condition that presenting the response of the external source is effective. It would then appear that, at least in the present situation, the variable of congruence of commitment is considerably stronger in its effect than that of opinion discrepancy.

In the situation of incongruent commitment *S* faces a somewhat different set of circumstances in that the two standards set by the external source are in considerably greater alignment than in the congruent condition. Both the standard represented by the response of the other person and the commitment standard are presented in a direction away from the *S*'s original position, although they are completely aligned only in the case of three-unit opinion discrepancy. Again assuming relative equality of response potential among persons for each standard, the mean amount of change obtained can be expected to increase as the amount of opinion discrepancy increases. Here, too, the results indicate an insignificant tendency toward a positive relationship. Incongruent attribution seems also to affect response more strongly than does opinion discrepancy.

Why the effect of incongruent commitment in the present experiment differs from that in the earlier one is not clear. It would appear that the amount of opinion discrepancy alone did not produce the limiting effect on incongruent commitment found in the earlier study, although it may be necessary to explore the effects of more extreme amounts of opinion discrepancy before a conclusive statement becomes possible.[5]

It was suggested that the amount of emotionality or ego involvement aroused by opinion topics may be a critical variable affecting response under incongruent commitment. This variable was not directly treated in the present study although items less likely to arouse such response than those of the earlier study were used. It remains possible that some interrelationship between this aspect of the opinion topics and the amount of opinion discrepancy produced the constriction on the effect of the incongruent commitment treatment in the earlier study.

An alternative context for interpreting the results pertaining to attribution of position may be available. Presenting situational cues that make a person's membership in a particular group salient favors responses similar to those learned and adopted in the context of that membership. Charters and Newcomb (1952) demonstrated that Catholics were more likely to express opinions reflective of orthodox Catholicism when such cues were present in an otherwise nondenominational context than when membership cues were absent. Kelley (1955) offered similar evidence indicating greater resistance to an attempt to change opinions of Catholic high school students when membership cues are present. In the present experiment, attribution of position can be regarded as having increased the salience of *S*'s previous opinion responses.

This interpretation is more immediately plausible in the case of congruent attribution than for incongruent attribution. In the latter instance, a response hierarchy may be assumed that includes several alternative responses in addition to the pre-experimental questionnaire

[5] An unpublished pilot study similar to the present one in most respects but presenting one unit, three units, and five units of opinion discrepancy to *S*s holding extreme positions yielded similar findings to those reported above.

response, among them one similar to that phrased in the incongruent attribution. The situational presentation of the attribution may then tend to elicit the response corresponding to this previously nonapparent portion of the response hierarchy. In group terms, it is as if one of a person's multiple membership groups is made salient rather than another. From a general S-R interpretation, the response that is obtained is the one associated most strongly with the cues in the particular situation, which is in turn dependent on the person's learning history.

SUMMARY

The present study was concerned with the effect of prior attribution of opinion to a person on a subsequent attempt to influence his opinion. It was hypothesized that an attribution congruent with the recipient's current opinion results in resistance to influence while an incongruent attribution facilitates opinion change. The effects were predicted on the basis of a process termed "external commitment."

An earlier study had suggested that limits on the effects of external commitment may depend on the amount of opinion discrepancy involved in the influence attempt. In the experiment reported, three degrees of opinion discrepancy were systematically presented with three external commitment treatments; congruent attribution, incongruent attribution, and a control or no commitment condition.

The results support the hypothesized effects of external commitment. Within the situation studied, external commitment has a stronger effect than opinion discrepancy, although results for a control condition offer partial support for Hovland and Pritzker's (1957) finding that the greater the opinion discrepancy, the greater is the amount of change obtained.

The process of external commitment was discussed as involving socially mediated rewards for consistent behavior. An alternative interpretation treated attribution of position in terms of salience of an earlier response tendency.

REFERENCES

CHARTERS, W. W., & NEWCOMB, T. M. Some attitudinal effects of experimentally increased salience of a membership group. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology.* (Rev. ed.) New York: Holt, 1952. Pp. 415–420.

DEUTSCH, M., & GERARD, H. B. A study of normative and informational social influences upon individual judgment. *J. abnorm. soc. Psychol.*, 1955, **51**, 629–636.

FISHER, S., & LUBIN, A. Distance as a determinant of influence in a two-person serial interaction situation. *J. abnorm. soc. Psychol.*, 1958, **56**, 230–238.

FISHER, S., RUBINSTEIN, I., & FREEMAN, R. W. Intertrial effects of immediate self-committal in a continuous social influence situation. *J. abnorm. soc. Psychol.*, 1956, **52**, 200–207.

GOLDBERG, S. C. Three situational determinants of conformity to social norms. *J. abnorm. soc. Psychol.*, 1954, **49**, 325–329.

HOVLAND, C. I., CAMPBELL, E. H., & BROCK, T. The effects of "commitment" on opinion change following communication. In C. I. Hovland (Ed.), *The order of presentation in persuasion.* New Haven: Yale Univer. Press, 1957.

HOVLAND, C. I., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion.* New Haven: Yale Univer. Press, 1953.

HOVLAND, C. I. & PRITZKER, H. A. Extent of opinion change as a function of amount of change advocated. *J. abnorm. soc. Psychol.*, 1957, **54**, 257–261.

KELLEY, H. H. Salience of membership and resistance to change of group-anchored attitudes. *Hum. Relat.*, 1955, **8**, 275–290.

OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity and the prediction of attitude change. *Psychol. Rev.*, 1955, **62**, 42–55.

POWELL, N. J. *Anatomy of public opinion.* New York: Prentice-Hall, 1951.

ROSENBAUM, M. E., & ZIMMERMAN, ISABEL M. The effect of external commitment on response to an attempt to change opinions. *Publ. opin. Quart.*, 1959, **23**, 247–254.

YEAGER, W. H. *Effective speaking for every occasion.* (2nd ed.) New York: Prentice-Hall, 1951.

(Received May 6, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 21-24

A CONSTANT ERROR APPROACH TO THE STUDY OF DIMENSIONS OF SOCIAL PERCEPTION[1]

LEON H. LEVY
Indiana University

AND ROBERT D. DUGAN
State Farm Insurance Companies

RATINGS may be treated either as descriptive data to be used in the further study of the people rated, or they may be treated as behavior in their own right to be studied for what they may reveal about those performing the ratings and the processes involved in this task. It is with this latter use of ratings that this study is concerned.

The observation that ratings of individuals on a number of different traits tend to correlate with each other to a degree that seems unwarranted has given rise over the years to such concepts as the "halo effect" and the "logical error." Thorndike (1920), in coining the term "halo effect," referred to it as a constant error in psychological rating resulting from "suffusing ratings of special features with a halo belonging to the individual as a whole." Guilford (1954) described the logical error, earlier observed by Newcomb (1931), as an instance of judges giving "similar ratings for traits that seem logically related in the minds of raters."

In each of these cases, the same individuals are rated on a number of different traits and the observed intercorrelations between these traits has been assumed to reflect a constant error. However, given the procedure and assuming that the intercorrelations do not represent the true state of affairs between the traits involved, it is impossible to separate the operation of the halo effect and the logical error from each other, or from that of yet another equally tenable explanation. This is that the correlations may arise because the traits rated are not independent of each other from the standpoint of the perceptual process involved in making the judgments, but rather represent specific instances of a more limited number of dimensions of judgment or social perception.

This latter consideration would suggest that from an analysis of the nature of the constant errors or intercorrelations found in trait ratings, information might be obtained concerning the nature of certain aspects of social perception. This approach has been followed, in effect, by Asch (1946) and many after him. However, in none of these studies has the investigator controlled for the operation of the so-called logical error. In fact, one might consider most of them as concerned with exploring the nature of what Bruner and Tagiuri (1954) so aptly refer to as the "naive, implicit 'theories' of personality people work with when they form impressions of others."

Whereas in previous studies demonstrating correlations between trait ratings, these relationships could be said to reflect the experience of the rater such that he has learned that Trait A implies Trait B (Hays, 1958), the present study is concerned with the question of whether certain relationships exist between trait ratings which *cannot* be attributed to such learning, but must rather reflect more basic dimensions of social perception. The entire conceptual structure of the projective approach to the understanding of perception and personality rests upon the assumption that perception is functionally related to certain states of the perceiver and, hence, would suggest the existence of such dimensions as perceptual counterparts of these states. This study represents a first attempt at determining their nature.

METHOD

One hundred introductory psychology students (52 men and 48 women) served as subjects (*Ss*) in the experiment. Each was seen individually and required to make a dichotomous trait judgment in connection with a particular bipolar trait for each of 225 photo-

[1] This investigation was supported in part by a research grant (M-2067) from the National Institute of Mental Health, of the National Institutes of Health, Public Service.

The authors wish to thank Jane A. Kelly for running the subjects of this study.

TABLE 1
INTERRELATIONS BETWEEN TRAITS

Traits	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1 Good–bad	50	48	41	49	44	42	50	33	29	13	29	00	07	−02
2 Safe–dangerous		40	34	42	29	33	18	22	13	12	06	−02	13	−01
3 Kind–cruel			23	23	26	32	16	31	22	22	04	10	11	−05
4 Reliable–unreliable				55	49	45	40	30	36	16	14	03	14	01
5 Honest–dishonest					45	42	37	38	13	12	06	−12	16	14
6 Stable–unstable						34	35	25	17	25	31	04	19	03
7 Friendly–unfriendly							31	36	19	21	05	07	18	−07
8 Ambitious–lazy								09	11	19	16	−13	06	−03
9 Generous–tight-fisted									19	24	23	15	18	04
10 Happy–unhappy										04	10	26	18	17
11 Thoughtful–self-centered											12	20	17	07
12 Attractive–unattractive												14	08	−06
13 Relaxed–tense													−13	00
14 Warm–cold														13
15 Unanxious–anxious														

Note.—An r of .19 is required for significance at the .05 level and an r of .25 for significance at the .01 level.

graphs presented to them one at a time. *S*s were instructed to make their judgments on the basis of their first impression of the person. The photographs were of white male applicants for admission to graduate training in psychology at Indiana University who, for one reason or another, never became students in the program. Each photograph was mounted on a 3 × 4 in. piece of poster board and was assigned at random to one of 15 decks. All photographs within a deck were judged on the same trait and had the two poles of this trait typed on a label attached below the picture. For all *S*s the same picture was always judged in connection with the same trait, but the order of presentation of pictures varied randomly from *S* to *S*. Thus, each *S* judged all 225 photographs but made only one judgment for each.

The 15 traits rated may be seen in Table 1. Traits 2, 11, 13 and 15 were typed in the reverse order from that presented in the table when they were presented along with their accompanying pictures to *S*s.

Each *S* was given a score for each trait which represented the number of photographs judged in the positive direction. Therefore scores for each trait could range from 0 through 15.

The method of the present study permits a clear distinction to be made between the operation of logical error and any of its kindred concepts, and perceptual dimensions. Since logical error results from inferring a certain amount of one trait from a certain amount of another trait in the same person, it cannot be invoked to account for correlations between traits obtained through the method described here where the two traits were rated not as characteristics of the same person, but rather as characteristics of different persons. For having judges rate one variable per photograph means that the judges cannot consider one variable or trait about a photograph to imply another variable or trait about another photograph. By the same reasoning, such correlations could not be assumed to represent the true relationship between the traits involved for the persons rated. Hence, any correlations found between traits in the present study can only be attributed to certain aspects of the judges or the perceptual processes involved, viz., the existence of certain fundamental dimensions of perception which are variously represented by the traits being judged.

RESULTS AND DISCUSSION

Pearson product-moment correlations were calculated between the 15 obtained scores and the resulting matrix was subjected to a factor analysis in which the Thurstone Multiple Group Method of extraction was used. The resulting factors were rotated graphically to orthogonal simple structure with communality estimates corrected by the Wherry Iterative Method (Wherry, 1949). The graphic rotation was performed using only the factor loadings and a visual criterion of simple structure rather than using the variable names and meanings. The table of intercorrelations and the rotated factor loadings are presented in Tables 1 and 2.

From the intercorrelation matrix it is obvious that relationships do exist between attribution tendencies for certain traits. If these ratings had all been performed on the same persons it would have been reasonable to account for many of the correlations in terms of synonymity or logical implication or true relationship between the two traits involved. In this instance these explanations are not tenable, and so it is reasonable to propose as an explanation the existence of certain dimensions of perception which are responsible for the correlations obtained.

Although we recognize the likelihood that

TABLE 2
ROTATED FACTOR LOADINGS

Traits	Factors				
	I	II	III	IV	h^2
1 Good–bad	.75	.32	.07	.06	.67
2 Safe–dangerous	.56	.37	.02	.05	.45
3 Kind–cruel	.45	.48	.07	.08	.44
4 Reliable–unreliable	.65	−.09	.36	.00	.56
5 Honest–dishonest	.65	.03	.30	−.15	.54
6 Stable–unstable	.62	−.02	.30	.14	.49
7 Friendly–unfriendly	.56	.12	.17	.05	.36
8 Ambitious–lazy	.49	.02	.17	.00	.27
9 Generous–tight-fisted	.44	.13	.11	.18	.25
10 Happy–unhappy	.33	.08	.08	.23	.17
11 Thoughtful–self-centered	.27	.04	.07	.34	.20
12 Attractive–unattractive	.25	.00	.05	.36	.19
13 Relaxed–tense	.00	.07	−.05	.49	.25
14 Warm–cold	.21	−.03	.10	−.04	.06
15 Unanxious–anxious	.02	.12	−.02	.01	.02

the factorial structure is dependent upon the limited number of traits sampled, we propose that the factors extracted from the matrix of intercorrelations are representative of certain major dimensions of social perception. From this standpoint, Factor I, which is a general factor, has been labeled General Evaluative Bias. This factor would seem to suggest that there is always a "halo effect" operating in that individuals differ in their tendency to react positively or negatively to another person and that this tendency is manifest throughout the realm of those traits sampled which permit an evaluative judgment to be made.

Factor II has been labeled Harmfulness, suggesting the importance of potential threat or hazard in the perception of others. Factor III seems most appropriately designated as Dependability in that a person described as reliable, honest, and stable, would be one that could be depended upon whereas the contrary would be true for a person described in opposite terms. Factor IV is somewhat more difficult to define, but has been tentatively labeled Affability.

We propose that these four factors, while not in any way exhausting the realm of possible dimensions of social perception, may be taken as representing four questions that each person asks about another person, and that there are individual differences in their tendencies to answer them in the affirmative or negative. These questions might be paraphrased as follows: for Factor I, "Is he good or bad?"; for Factor II, "Is he likely to cause me any trouble?"; for Factor III, "Can I count on him?"; and for Factor IV, "Would I want him as a friend?"

For the personality theorist the present study might be taken as demonstrating that, for each of the dimensions isolated, persons vary in their adaptation level (Helson, 1955), or point of subjective equality, for their judgments of others. This interpretation immediately raises the question of the personality variables of which this parameter may be a function as well as that of those behaviors which may be expected to vary as a consequence of variations in it. Rather than speculate on the answers to these questions at this point, we simply raise them to indicate the directions for further research suggested by the present study.

In contrast to the findings of Asch (1946) and those of others following him, the present study gives no support for the centrality of the "warm–cold" dimension. Its only significant correlation is with "stable–unstable," and it fails to have a significant loading on any of the factors extracted. Since the distribution of scores on this dimension had a range from 4 to 12, the finding cannot be attributed to restricted variance and must therefore find its explanation elsewhere. In order to determine whether the unreliability of this dimension precluded its correlating with other dimensions, each *S*'s scores for his odd and even numbered pictures were correlated. The resulting correlation, corrected by the Spearman-Brown formula, was .32 suggesting that reliability, while not high, cannot be held entirely accountable for the lack of correlation found. One possible explanation for the discrepancy may reside in the difference in procedures used in this study and in previous ones. It may be that different processes were tapped in each. Thus, in studies on the Asch model *S* was always supplied with information regarding this dimension, such studies could only reflect the possible personality-perceptual bases responsible for relationships between "warm–cold" and other dimensions. In the present study, on the other hand, relationships between dimensions or traits can only occur through relationships in those personality-perceptual

processes responsible for differential tendencies in ascribing them to people in a context of minimal information.

The complete absence of correlation between "unanxious–anxious" and any of the other traits is somewhat more puzzling. Again, from inspection of the distribution of scores for this trait, this finding cannot be explained in terms of restricted variance, and the odd–even reliability, corrected by the Spearman-Brown formula, was in this case .37. It may be that some *S*s interpreted anxiety as eagerness rather than dysphoria, or it may be that if additional traits were included anxiety would emerge as part of another major dimension. Until further research is carried out these questions must remain unanswered.

The immediate practical implications of the present findings reside in the use of rating scales in the description or evaluation of persons. To the operation of the halo effect and the logical error, there must now be added the dimensions of social perception operative when ratings are made and which may also be responsible for some portion of the correlations found between ratings. Two ratings may correlate simply because they represent the same perceptual dimension. This is of particular importance where correlations between trait-ratings form the bases for discussions of personality types and syndromes (e.g., Sheldon & Stevens, 1942; Lorr, Rubinstein, & Jenkins, 1953; Cattell, 1957). Until it has been demonstrated that the judges are capable of producing ratings which are independent of each other, the personality types described by this procedure may be expected to reflect the perceptual processes of the judges as much as they do the nature of the objects of their judgments.

SUMMARY

A method was described for use in the study of dimensions of social perception which consisted of the factor analysis of intercorrelations between trait judgments of photographs where each judgment was made of a different stimulus. The method precludes the possibility that correlations between judgments represent learned contingencies and therefore may be taken as a means of isolating basic social perceptual dimensions that in turn may represent the perceptual counterparts of certain personality parameters of the judges.

Four factors were extracted from an intercorrelation matrix representing judgments on 15 bipolar traits. These were labeled as follows: Factor I—General evaluative bias; Factor II—Harmfulness; Factor III—Dependability; Factor IV—Affability.

The relationship between these findings and those of previous studies in social perception was discussed briefly, as was their implications for research using personality ratings in the description of personality types and syndromes.

REFERENCES

ASCH, S. E. Forming impressions of personality. *J. abnorm. soc. Psychol.*, 1946, **41**, 258–290.

BRUNER, J., & TAGIURI, R. The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. 2. Cambridge, Mass.: Addison-Wesley. 1954. Pp. 634–654.

CATTELL, R. B. *Personality and motivation structure and measurement*. Yonkers, N. Y.: World Book: 1957.

GUILFORD, J. P. *Psychometric methods*. (2nd ed.) New York: McGraw-Hill, 1954.

HAYS, W. L. An approach to the study of trait implication and trait similarity. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford Univer. Press, 1958. Pp. 289–299.

HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1955, **55**, 297–313.

LORR, M., RUBINSTEIN, E., & JENKINS, R. L. A factor analysis of personality ratings of outpatients in psychotherapy. *J. abnorm. soc. Psychol.*, 1953, **48**, 511–514.

NEWCOMB, T. An experiment designed to test the validity of a rating technique. *J. appl. Psychol.*, 1931, **22**, 279–289.

SHELDON, W. H., & STEVENS, S. S. *The varieties of temperament*. New York: Harper, 1942.

THORNDIKE, E. L. A constant error in psychological rating. *J. appl. Psychol.*, 1920, **4**, 25–29.

WHERRY, R. J. A new iterative method for correcting erroneous communality estimates in factor analysis. *Psychometrika*, 1949, **14**, 231–241.

(Received April 17, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 25-30

A COMPARISON OF SEMANTIC STRUCTURES IN AMERICAN SOUTHWEST CULTURE GROUPS[1]

GEORGE J. SUCI[2]

Gerontology Branch, National Heart Institute, and National Institute of Mental Health

AN OBJECT is perceived, or a concept described, in terms of a set of attributes, and most attributes, but not necessarily all, have linguistic equivalents that permit their communication. This assumption is, in part, the basis for an operational definition of meaning—the semantic differential, extensively described elsewhere (Osgood, Suci, & Tannenbaum; 1957). Briefly, concepts or objects are judged on rating scales defined with adjective opposites (e.g., *good–bad*). In English, adjectives appear to be related such that only a few dimensions are needed to account for most variation in such judgments. These dimensions define a semantic space—and the meaning of a concept is given by the coordinates of its location in this space.

A concept has the same meaning for two individuals or two culture groups when the concept is located in the same position of the respective semantic spaces. That such an equivalence between concepts can be said to exist depends on the equivalence of the semantic spaces; i.e., the dimensions, and therefore attributes, must mean the same for the two individuals. To avoid circularity, the meaning of attributes cannot be inferred from the meanings, or locations, of the concepts; instead, such meaning is inferred from the totality of relations among the attributes themselves. If the same relations exist between a set of attributes for two different individuals, their semantic spaces are equivalent and their meanings of a concept may be compared. Obviously without such equivalence of space, communication between individuals regarding the meanings of their concepts is seriously impeded.

Extremely different subjects have shown highly similar semantic structures. Suci (1952) found that a group of high and a group of low scorers on the California F Scale judged ethnic concepts in similar semantic frames of reference. The meanings of the ethnic concepts, however, differed significantly between the groups. The same kind of results were demonstrated when Republican and Democratic subjects rated politically relevant concepts (Osgood et al., 1957, pp. 104–124). Meanings of concepts like FDR, MCCARTHY, POLICY IN CHINA differed significantly while the attribute relations remained constant. Similar results are reported by Bopp (1955) for normal and schizophrenic subjects.

Does semantic constancy occur across subjects selected from different cultural and linguistic backgrounds? Kumata and Schramm (1956) showed equivalent semantic structures for Korean bilingual exchange students, Japanese bilingual exchange students, and American natives when the semantic judgments were made in the respective languages. Later Kumata (1957) showed the same kind of equivalence for Japanese and American monolinguals, thus extending the generality of the Kumata and Schramm findings. Triandis and Osgood (1958), aiming more directly at a test of Whorf's hypothesis (1956) that language

[1] This study was conducted as a part of the Southwest Project in Comparative Psycholinguistics, sponsored by the Committee on Linguistics and Psychology of the Social Science Research Council under a grant from the Carnegie Corporation of New York. The data were analyzed at the University of Illinois Computer Laboratories. For a more complete description of the entire project, see Casagrande (1956). The author wishes to express his appreciation to the following members of the project who were kind enough to collect the data used in this study: Susan Ervin for Navaho; J. B. Carroll, Lois Elliott, and H. S. Maclay for Hopi; J. R. Bastian, Sally Bastian, E. H. Lenneberg, and S. S. Newman for Zuni; and S. Saporta for Spanish.

[2] This study was carried out when the author was at the University of Illinois.

determines cognition, found that Greek and American monolinguals used similar semantic spaces. Thus, Triandis and Osgood, as well as Kumata, have produced evidence against the Whorfian hypothesis, at least with respect to meaning as measured by the semantic differential. The present study extends the comparison of semantic structures to subjects from the American Southwest—Spanish, Hopi, Zuni, and Navaho subjects.

METHOD

In general, the design consisted of constructing a semantic differential and administering it to subjects (*Ss*) from each culture group, intercorrelating the semantic scales for each group, factoring each of these sets of relationships, and comparing the obtained factor structures. The Southwest groups will be designated as "experimental" groups to differentiate them from an English-speaking "control" group which is described below.

The Semantic Differential

A first selection of bipolar adjective scales was made as follows. Each of 42 nouns was translated into the native language by an informant from each experimental group. The nouns were read one at a time as stimuli to each informant who was instructed to respond with as many appropriate adjective forms as he could. Each "adjective" obtained in this fashion was read back to the informant who was instructed to respond with a proper opposite. Of the total number of bipolar opposites obtained in this fashion, 25 were common to three of the experimental groups. These were retained as items in the semantic differentials given to the *Ss*. In the field it was necessary for each investigator to adjust his list because of poor translations, or *Ss*' refusals to respond to some of the items. This resulted in the use of different polar terms by different investigators, and, in the final analysis, only 15 scales were common to all groups.

Ten concepts were selected to be rated on the scales. As with the scales, many of the concepts were found unsatisfactory because of translation difficulties or because they touched on sensitive areas and threatened rapport with the *Ss*; therefore, new concepts had to be substituted in some cases and different sets of concepts resulted for some groups. *Ss* of all groups rated: MEXICAN, CORN, MAN, WOMAN, HORSE, and RAIN. The Zuni group also rated COYOTE, RED, YELLOW, and BLUE-GREEN. The Hopi and Navaho also rated NAVAHO, WHITE PEOPLE, COYOTE, and SELF, and the Spanish group rated RELIGION, DEATH, ARMY, and GODFATHER.

Control for Concept Differences

Semantic structures may vary with different concepts (Osgood et al., 1957, pp. 176–188); therefore, it was necessary to estimate the extent of semantic space variation due to differences in concepts rated by each group. All the different concepts were presented to a "control" group of English speakers, monolingual college students at the University of Illinois. Three semantic spaces were factored out, each one based on a subset of concepts rated by an experimental group. The three semantic spaces were compared. Since each English speaking *S* was his own control, differences between these structures could be attributed only to differences in concepts, not to differences in language or culture; therefore, factor differences between the control semantic structures implied that comparisons between experimental groups were not valid for those factors. In the analysis, no comparisons between experimental groups were made when such differences occurred.

Subjects

The experimental *Ss* were 26 Zuni, 28 Hopi, and 27 Navaho Indians, and 32 Spanish speakers from the Albuquerque, New Mexico area. All *Ss* were fluent speakers of their native language, but English was spoken with varying degrees of competence. The Spanish, Hopi, and Zuni *Ss* were bilingual but dominant in their own language as judged by the linguists working with these groups. The extent of bilingualism in the Navaho *Ss* was judged on the basis of a bilingualism test.[3] Of the 27 Navaho, 14 were English dominants, 6 were Navaho dominants, and 7 were Navaho monolinguals. The median age and age-range for each group was as follows: Zuni, 35, 12–70; Hopi, 24, 12–64; Navaho, 45, 17–75; Spanish, 26, 18–34; English speakers, 21, 19–26. The sex distribution, male and female respectively, in each group was as follows: Zuni, 15, 11; Hopi, 12, 16; Navaho, 8, 19; Spanish, 21, 11; American, 18, 15.

Instructions and Procedure

The literate Spanish and control groups received standard instructions (Osgood et al., 1957, pp. 82–84). Administration was in a group using paper-and-pencil forms printed in the respective native languages.

Since the Indian people are nonliterate in their native languages, *Ss* from these groups could not respond to the standard rating procedure and a modified technique had to be devised. Twenty-five lines, one for each scale, with the ends and the middle of each line marked, were mimeographed on legal size paper. The lines were six inches long. The instructions, given below, will explain the procedure employed in the administration of the items to the Indian *Ss*.

> We would like to know how you describe or talk about certain things. For example, you may think of something being very fast or very slow, or a little bit strong or a little bit weak. Suppose we wanted

[3] The bilingualism test was developed by Susan Ervin.

TABLE 1

FIRST THREE FACTORS AFTER BIQUARTIMIN ROTATION

Scales	Navaho			Hopi			Zuni			Spanish			A-N-H[a]			A-Z[b]			A-S[c]		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
heavy–light	07	47	−01	09	45	−29	11	10	53	−35	34	07	−04	47	33	01	12	02	−21	49	−12
hard–soft	−26	26	−05	−01	30	−25	−08	23	−08	−28	22	26	−29	28	−10	−38	45	−05	−07	30	−18
pretty–ugly	51	07	−02	72	01	−14	64	−06	00	79	00	−01	61	−06	10	66	01	06	41	−03	27
hot–cold	−40	40	08	20	−01	−39	20	−16	42	42	14	38	01	00	54	−05	02	59	−08	05	65
clean–dirty	34	26	00	58	−07	14	65	03	10	61	08	−12	75	−05	−12	72	03	−12	67	−06	−10
dry–wet	−12	−15	28	−05	−22	−43	12	−18	−15	−13	−11	23	−20	01	15	−29	22	37	−09	−01	17
industrious–lazy	41	01	24	48	05	07	16	67	−02	15	62	−06	34	32	−04	17	66	07	29	46	17
happy–sad	24	42	−13	53	07	02	59	08	−02	53	11	08	39	−06	31	37	00	46	19	03	53
rich–poor	00	49	−06	54	03	07	60	11	15	33	11	−43	61	24	−15	59	26	−07	61	16	−18
sweet–sour	32	02	02	21	−08	47	33	24	17	78	−03	06	58	−20	15	63	−16	10	44	−22	29
long–short	05	48	03	13	33	08	23	31	21	−04	51	−01	13	54	−03	14	36	24	07	43	−30
fast–slow	03	−03	65	−02	53	01	−10	71	12	08	45	−10	08	60	−10	−03	63	−07	09	48	08
strong–weak	12	46	−07	01	64	05	02	72	−04	−03	64	00	01	58	16	−04	54	11	02	64	10
straight–crooked	05	05	00	06	−19	05	50	05	−05	38	20	−05	53	17	21	55	12	09	48	08	04
good–bad	53	18	01	42	05	34	31	46	−26	77	02	−01	69	02	12	71	−02	00	52	06	25
% Common Variance	14.9	17.3	7.1	25.1	15.4	11.4	28.8	26.4	8.9	41.6	19.9	6.4	35.9	19.6	8.7	36.1	20.3	10.1	22.1	17.2	13.8

Note.—Decimal points are omitted.

[a] American control for Navaho and Hopi concepts.

[b] American control for Zuni concepts.

[c] American control for Spanish concepts.

to know how strong or weak something like MOUNTAIN is. You could think of this end (experimenter [*E*] points to one end of a line) as being something very, very strong and this end as something very, very weak. Now since MOUNTAIN is very, very strong, it would be very much like the very, very strong thing on this end and you would show me, by pointing (or pencil mark), that this is where MOUNTAIN belongs.

On the other hand, something like GRASS is not strong, but quite weak, and you would show me this by placing it close to this thing which is very, very weak. A TREE is quite strong but certainly not as strong as MOUNTAIN, so you might put TREE here, on the strong side, but not as close to the very, very strong end as MOUNTAIN. A BUSH is not as strong as a TREE, but neither as weak as GRASS, you may put it between GRASS and the TREE, here, but a little on the strong side. (The *E* repeats the above with BREAD, BLANKET, SPIDER, and SNAKE, in that order form good to bad, on scale *good–bad*.)

Now think of this end as very, very active and this other end as very, very passive. Now where do you think MOUNTAIN should go? COW? DEER? BEE?

Now here are a number of lines. I will tell you for each line what each end is. Then I will ask you to put something like RAIN where you think it belongs on the line. There is only one place on the line where you cannot put the thing you are to describe, and that is on this point (*E* points to center). You may put it close to this point, but not on it (this is to help force the *S* to make a discrimination).

The above instructions were followed as closely as possible by each investigator. One concept was presented with each sheet of scales. The ends of the scale and the concept were read to the *S* for each judgment. The items were given in the native language to all *S*s. English instructions were used except with the Navaho who were instructed in Navaho.

RESULTS

Digits 1 through 7 were assigned in left to right order to the seven points on the scales of the English and Spanish semantic differentials. A score was the digit corresponding to the check-mark representing the *S*'s rating. The Indian differentials were scored by measuring the distance, in one-tenth inches, from the left side of the scale to the check-mark locating the concept. Each *S*'s rating formed a scale by concept matrix of order 15 × 10. Since

TABLE 2

COEFFICIENTS OF FACTORIAL SIMILARITY

Factor 1								Factor 2						
N	H	Z	S	N-H[a]	Z[b]	S[c]		N	H	Z	S	N-H[a]	Z[b]	S[c]
1.00	.72	.61	.69	.79	.79	.72	Navaho	1.00	.60	.42	.65	.60	.45	.63
	1.00	.91	.81	.87	.83	.82	Hopi		1.00	.74	.79	.87	.70	.88
		1.00	.83	.91	.88	.86	Zuni			1.00	.84	.76	.80	.79
			1.00	.90	.90	.86	Spanish				1.00	.92	.89	.95
				1.00	.99	.96	American-Navaho-Hopi					1.00	.88	.97
					1.00	.94	American-Zuni						1.00	.90
						1.00	American-Spanish							1.00

Factor 3								Factor 4						
N	H	Z	S	N-H[a]	Z[b]	S[c]		N	H	Z	S	N-H[a]	Z[b]	S[c]
1.00	−.15	.11	.03	−.05	.08	.13	Navaho	1.00	.64	.02	−.23	−.15	.18	−.13
	1.00	−.27	−.55	−.32	−.42	−.09	Hopi		1.00	−.24	−.32	−.50	.23	−.28
		1.00	.12	.45	.17	.08	Zuni			1.00	.71	.59	−.17	.66
			1.00	.61	.60	.47	Spanish				1.00	.60	.09	.78
				1.00	.80	.73	American-Navaho-Hopi					1.00	−.16	.86
					1.00	.88	American-Zuni						1.00	−.04
						1.00	American-Spanish							1.00

[a] American control for Navaho and Hopi concepts.

[b] American control for Zuni concepts.

[c] American control for Spanish concepts.

there were N such matrices, a $15 \times 10N$ matrix was formed for each group.

Product-moment correlation coefficients were calculated between scales for each group by summing the cross-products over both *S*s and concepts so that the number of pairs entering into each coefficient was $10N$. The 15×15 correlation matrix for each group was factored using the principal-axis technique with a communality estimate in the main diagonal, and extracting all 15 factors.[4] The first four principal components were rotated into simple structure using a modification of Carroll's method (1953).[5] Finally, to measure the similarity between the factors, indices of similarity (Wrigley & Neuhaus, 1955) were calculated between related factors.

As shown in Table 2, the first three factors for the control groups were highly similar, but the fourth factors were quite dissimilar. Such dissimilarity implied that differences on Factor 4 could be attributed to concept differences, and that inferences regarding language or culture differences on this meaning dimension could not be made.

The coefficients of similarity for Factors 1 and 2 were consistently positive for all the groups. Factor 3, however, produced some extremely low, and even negative, similarity coefficients. Triandis and Osgood (1958) accepted a value of .75 as the lower limit for statistically significant factorial similarity. Not all coefficients exceeded this value for Factors 1 and 2. The Navaho group consistently produced coefficients less than .75. The other groups, with the exception of Zuni and Hopi whose coefficient for Factor 2 had a border line value of .74, were above this value for Factors 1 and

[4] The tables of intercorrelations and the first four principal components for each group have been deposited with the American Documentation Institute. Order Document No. 6238, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[5] The rotations were performed by J. B. Carroll using an IBM Type 650 program developed by J. B. Carroll, F. Weinfeld, and A. Beaton.

2. None of the experimental groups showed relations above .75 for Factor 3.

From the loadings in Table 1, the first factor was readily interpretable as evaluative with scales like *pretty–ugly*, *clean–dirty* (with the exception of Navaho) and *good–bad* (with the exception of Zuni) producing some of the higher loadings for all the groups. This factor explained the most variance for every group except the Navaho where it accounted for only 14.9%, slightly less than the 17.3% shown by the Navaho second factor. The second factor appeared as a combination of potency (*strong–weak*) and activity (*fast–slow*) for all groups except Navaho who displayed mostly potency scales on this factor with *fast–slow* appearing independently on the third factor. This combination of activity and potency has been termed "dynamism" (Osgood et al., 1957, p. 74) and commonly appears with English speakers. The third factor was not readily interpretable for any group.

DISCUSSION

With the exception of the Navaho group, a high degree of similarity in the semantic structures of *S*s from different cultural backgrounds was found. Since most *S*s were bilinguals, the results do not have the generality of those from the Triandis and Osgood (1958) or the Kumata (1957) studies based on monolingual subjects.

Conclusions are further limited in that only a small amount of the judgmental variance is explained in some cases, especially the Navaho. The first three Navaho factors accounted for only 39% of the reliable variance. In contrast, the three Spanish factors accounted for 68% which concurs with the 70% accounted for by the Greek and English factors (Triandis & Osgood, 1958). In other words, a vast amount of reliable attribute space remains undefined for the Navaho group. To a lesser extent this is also true for the Hopi, whose first three factors account for about 52%. Thus, for the Navaho especially, only a narrow base for comparison with other groups was available. Only further research can tell if this lack of accountable variance explains the Navaho deviation from the other groups. Such limitations, although important and suggestive for further research, should not be permitted to hide the significance of the high degree of similarity. Despite differences in culture, concepts, method of administration and education, similarity between groups was high. Further, the main two dimensions were interpretable as evaluative, and activity-potency, or dynamism, the same factors found with all the other culture groups studied. Thus, taking into account other studies in this series, a high degree of similarity may be assumed to exist in the semantic frames of reference used by subjects of different cultural background.

SUMMARY

The present study is one of a series investigating the similarity of semantic structures for different groups of subjects—in this case *S*s from different American Southwest cultures. The factor structures of a sample of semantic scales indicated that Zuni, Hopi, Spanish, and English-speaking *S*s define a semantic space with similar evaluative and dynamism dimensions. Although the semantic space for a group of Navaho could be defined with the same dimensions, the similarity measures were consistently lower for this group. A third factor for each group could not be interpreted and was highly dissimilar for all groups. Since most *S*s were bilinguals, the generality of the findings was limited in comparison to findings from studies using only monolinguals. Only a small amount of variance is accounted for by the first three factors for some groups, especially the Navaho, suggesting that a large amount of attribute space needs to be defined with further research.

REFERENCES

BOPP, JOAN. A quantitative semantic analysis of word association in schizophrenia. Unpublished doctoral dissertation, Univer. Illinois, 1955.

CARROLL, J. B. An analytic solution for approximating simple structure in factor analysis. *Psychometrika*, 1953, **18**, 23–38.

CASAGRANDE, J. B. The Southwest project in comparative psycholinguistics: A progress report. Soc. Sci. Res. Council, *Items*, 1956, **10**, 41–45.

KUMATA, H. A factor analytic investigation of the generality of semantic structure across two selected

cultures. Unpublished doctoral dissertation, Univer. Illinois, 1957.

KUMATA, H., & SCHRAMM, W. A pilot study of cross-cultural methodology. *Publ. opin. Quart.*, 1956, **20**, 229–237.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

SUCI, G. J. A multidimensional analysis of social attitudes with special reference to ethnocentrism. Unpublished doctoral dissertation, Univer. Illinois, 1952.

TRIANDIS, H. C., & OSGOOD, C. E. A comparative factorial analysis of semantic structures in monolingual Greek and American college students. *J. abnorm. soc. Psychol.*, 1958, **57**, 187–196.

WHORF, B. L. *Language, thought and reality: Selected writings of Benjamin Lee Whorf.* New York: Wiley, 1956.

WRIGLEY, C. F., & NEUHAUS, J. O. The matching of two sets of factors. *Amer. Psychologist* 1955, **10**, 418–419.

(Received May 20, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 31-38

PATTERNS OF ROLE DOMINANCE AND CONFLICT IN PARENTS OF SCHIZOPHRENIC PATIENTS[1]

AMERIGO FARINA
Duke University

A GREAT deal of effort has been directed toward determining the possible role played by familial factors in the development of schizophrenia. In past years the role of the mother of the schizophrenic patient has frequently been given primacy (Rosen, 1953; Fromm-Reichman, 1948). However, the nature of the total family structure and the pattern of parental relationships have also received attention, especially in more recent reports. According to these studies, a characteristic interaction pattern of parents of schizophrenic patients is that of greater dominance of one over the other than would be expected in a population of parents of nonschizophrenic offspring (Arieti, 1955; Ellison & Hamilton, 1949; Gerard & Siegel, 1950; Kohn & Clausen, 1958; Lidz, Cornelison, Fleck, & Terry, 1957a; Reichard & Tillman, 1950). Also some preliminary studies have indicated a systematic relationship between the sex of the dominant parent and important patterns of adjustment on the part of the son. In several instances (Rodnick & Garmezy, 1957), it has been found that patients from father-dominated families tend to be married, have numerous friends of both sexes, and generally have made a good premorbid adjustment. Patients from mother-dominated families, on the other hand, tend to be single, social isolates, and to have made a marginal social-sexual adjustment prior to psychosis. These trends are supported by Opler's (1957) finding that patients from mother-dominated subcultural groups tend to be pallidly asexual, whereas those from father-dominated subcultures tend to be more active sexually. Parents of schizophrenic patients, in addition, are reported to relate to each other in a manner characterized by heightened conflict and hostility (Frazee, 1953; Friedlander, 1945; Gerard & Siegel, 1950; Lidz, Cornelison, Fleck, & Terry, 1957b; Tietze, 1949).

This study was primarily designed to investigate the reported association between the sex of the dominant parent and the premorbid adjustment of the schizophrenic son. In addition, the reported conflict in the interaction of these parents as well as the inequality of the dominance relationship seemed to merit investigation. Considerable difficulty is encountered, however, in determining what is usually meant by dominance and conflict in this area of investigation. Many of the studies cited are primarily clinical-qualitative reports and little or no attempt has been made to define these two variables. None of these studies had attempted to measure systematically any aspect of actual parental interaction. In spite of this difficulty, the hope in the present study was to establish operationally defined indices of dominance and conflict which would do minimal violence to commonly held connotations of these terms. Thus, in a standardized situation where the parents are interacting, the more dominant partner can be defined as the one who speaks the most, whose decisions are accepted by the other, etc. In the same situation, extent of conflict can be measured by number of interruptions by each parent, failure to reach agreement, etc. Observing the actual interactions between the parents should have a decided advantage over drawing inferences from paper-and-pencil scales completed by each. If, in addition, the standardized situation were so structured that the parents were unaware of what was being measured, their understandable defensiveness about revealing patterns of conflict and dominance in their relationship could be minimized.

Other studies which have used comparable measures of dominance and conflict give some indication of their usefulness. Matarazzo,

[1] This report is based on a PhD thesis submitted to the Graduate School of Arts and Sciences at Duke University. The research was in part financed by a research grant (M-629) from the National Institute of Mental Health. Grateful appreciation is expressed to Norman Garmezy, Eliot H. Rodnick, and Edward E. Jones for their help and suggestions. Thanks are due to the staffs of the State Hospital of Butner, North Carolina and Oteen VA Hospital, Oteen, North Carolina for their help in obtaining *Ss*.

TABLE 1
CHARACTERISTICS OF PARENT GROUPS

	Controls	Goods	Poors
Age of Father			
Mean	59.7	55.8	60.5
SD	6.9	9.1	10.2
Age of Mother			
Mean	56.6	52.3	56.7
SD	8.1	8.2	9.7
Years Married			
Mean	37.4	32.0	34.8
SD	7.2	7.0	9.5
Grades Completed by Mother			
Mean	6.4	8.1	8.6
SD	2.8	2.6	3.9
Grades Completed by Father			
Mean	7.3	7.2	9.5
SD	3.0	3.3	3.7
Number of Children			
Mean	5.2	4.2	4.3
SD	2.6	1.8	2.5
Number of Male Children			
Mean	3.4	2.6	2.2
SD	1.4	1.2	1.1

Saslow, and Guze (1956) have demonstrated that somewhat similar indices have excellent inter-rater reliability and have considerable stability over time for a given individual. Strodtbeck (1955) and Cervin (1957) have employed similar measures for identifying the dominant member of a dyad and their results indicate considerable usefulness for interaction situation techniques.

METHOD

Subjects. The *S*s were 36 pairs of white parents who were divided into three groups of 12 couples each. The grouping of these parents was determined by the premorbid adjustment and type of illness of their biological sons. Parents in the control group had sons hospitalized for pulmonary tuberculosis with a medical history free from psychiatric illness or disturbance. These parents were selected as control *S*s because tuberculosis shares a number of important features with mental disorder. Society is antipathetic to both illnesses, and both result in prolonged hospitalization and social isolation of the patient. Moreover, it was possible to employ an identical procedure with these parents and the parents of schizophrenic patients. Parents in the Good Premorbid group (Goods) had sons hospitalized for schizophrenia with the symptom picture uncomplicated by organic factors or mental retardation. In addition, these sons had made a good adjustment prior to psychosis as indicated by a Phillips Premorbid Scale score of 15 or less (Phillips, 1953). Parents in the Poor Premorbid group (Poors) were selected by the same criteria as the Goods with the difference that their sons had made a poor adjustment prior to psychosis as indicated by a Phillips Premorbid Scale score of 16 or more. The validity of this dichotomy has been reported in detail elsewhere (Rodnick and Garmezy, 1957). The Good sons were mostly married and all had numerous friends, whereas, the Poors were all single and, typically, had been social isolates. For all groups, both parents and the hospitalized son had lived together as a family unit during the son's childhood and adolescence, and the parents were living together at the time of the study. The mean age of all sons was 28.7 with a range extending from 19.5 to 36 years. None had been hospitalized for his present illness prior to the age of 17. The groups of sons were quite comparable with respect to age and education. In addition, the Tuberculous and Good groups were nearly identical with respect to total years spent in the hospital and number of years since first hospitalization. However, the Poors tended to have been hospitalized at a younger age and had spent significantly more time in the hospital in comparison with the other two groups.

Pertinent data on the groups of *S*s (couples) are presented in Table 1. The only statistically significant differences among these groups for the characteristics appearing in Table 1 occur for number of male children. The control parents had significantly more male children than the other two groups. Although it is difficult to classify many of these parents into usual socioeconomic class categories because many were engaged primarily in farming, the total sample might best be described as lower middle class, and the groups appear comparable in this respect. With the exception of one mother, all *S*s are native Southerners.

Procedure. When a set of parents satisfied the criteria listed above, they were sent a letter from the Clinical Director or Director of Professional Services of the hospital in which their son was a patient. This was the standard method of communicating with relatives of patients, and most parents had previously been contacted in this manner by the hospital. The letter stated that *E*, a clinical psychologist, was collecting information about patients hospitalized there and that the information was considered of importance in understanding and treating the son. The recipients of the letter were asked to make an appointment with *E* at their son's hospital.

When the parents arrived, they were told that information about their son and family was desirable and might help their offspring.[2] They were then interviewed about family composition and the psychotic son's premorbid history. After this interview, the father was taken to another office and asked to complete a modified form of the Parental Attitude Research Inventory (PARI) (Schaefer & Bell, 1955). *E* then returned to the office in which the mother was waiting and explained that the subsequent part of the interview would be recorded to insure *E*'s recall of what was said. *S* was then given the following instructions:

> Mrs.—, here's what I'd like us to do now. We are interested in knowing how a mother handles problems that come up with children when her

[2] The data collected actually served a clinical purpose. Such data were made available to professional personnel who were working with sons of these parents.

husband is not around. I'm going to describe some problems that come up more or less often, and I'd like you to tell me what you would do. Remember, imagine that your husband is not around to help you so that you have to handle these problems yourself.

Twelve hypothetical problem situations, adapted from a study by Jackson (1956), were then read, and *S* was given unlimited time to respond to each one. Three examples of these problem situations are given below:

1. Your twelve-year-old son gets home from school. He throws his books down, telling you what you have been teaching him is wrong and that he is not going to listen to you any more.

2. A gang of boys calls to your eight-year-old son to come out and play. You don't think it's good for your son to play with these boys, but now he starts to leave the house to go with them.

3. Your seventeen-year-old son has a chance to take a job which you are sure would be good for him. He knows you would like him to take it, but he doesn't want it. Instead, he wants to take a job which you think would not be good for him.

When the task was completed, the roles were reversed. The mother was taken to the other office to complete the PARI, and the father was given the same instructions and the same hypothetical situations. Finally, both parents were once more brought together in the office containing the tape recorder and given the following instructions:

Both of you have told me how you would handle problems that come up with children when you are by yourselves. We are also interested in knowing how parents would handle these problems together. Imagine that these same problems came up when both of you are there. I would like both of you to agree about what you would do and tell me what that is. Be sure both of you say "agreed" when you are finished.

The tape recorder was started once more and the problem situations were again read by *E*. For all *S*s, these hypothetical problems were read in the same order.

Throughout the interaction between the parents, great care was exercised to insure that *E* did not influence their behavior. Verbal reinforcement was given only once—and this came in all instances—at the end of the parent's first response during the initial interview in which each was seen individually. Furthermore, the extent and quality of *E*'s participation in the interaction was guided by the Jackson content scoring categories (Jackson, 1955, 1956). All questions by *E* were designed to elicit scorable responses from the parents or to determine whether or not agreement had been reached between them.

Indices of Dominance. The dominance and conflict measures employed in the analyses were taken exclusively from the interview with both parents with the exception of the "Yielding" indices, which were derived from both the individual and joint sessions. The several indices of dominance were operationally defined as follows:

1. Speaks First—The number of times one parent spoke first in 12 situations minus the number of times the other spoke first.

2. Speaks Last—The number of times one parent spoke last in 12 situations minus the number of times the other spoke last.

3. Total—First and Last—Sum of 1 and 2 above.

4. Passive Acceptance of Solution—The number of times in 12 situations one parent passively accepted the other's solution minus the number of times the other behaved similarly. A simple unelaborated agreement in response to the spouse's solution constituted passive acceptance.

5. Total Time Spoken—One parent's total speaking time for 12 situations relative to the other's total speaking time.

6. Yielding-Maximum—The content of the responses to the problem situations was scored along the dimension of amount of pressure to conform which the parent, by his or her action, would bring to bear on the child. An ordinal scale developed by Jackson (1955) was employed for this purpose, and yielding scores were defined in terms of scale points yielded by one parent in moving toward an acceptance of the spouse's solution. For example, if the mother was very lenient, whereas the father applied great pressure to the child in the individual interviews, and together they arrived at a lenient solution, then the father was considered to have yielded to the mother. When more than one solution was offered to one problem situation, Yielding-Maximum was always based on the most severe solution offered by the parents, both in the individual and joint interviews.

7. Yielding-Minimum is identical to Yielding-Maximum except that it was always derived from the least severe solution offered.

Indices of Conflict.[3] These indices were chosen on the assumption that they would reflect the degree of respect and tolerance the *S*s had for each other and would indicate the degree to which they could compromise and cooperate. They were operationally defined as follows:

1. Frequency of Simultaneous Speech—The number of occasions in the interview during which both parents spoke concurrently.

2. Duration of Simultaneous Speech—The total time in seconds during which both parents spoke concurrently.

3. Interruptions - Mother—The number of times the mother interrupted the father during the interview.

4. Interruptions - Father—The number of times the father interrupted the mother.

5. Interruptions - Total—Sum of 3 and 4 above.

6. Disagreements and Aggressions - Mother—The number of times the mother disagreed with or aggressed against the father during the interview.

7. Disagreements and Aggressions - Father—The number of times the father disagreed with or aggressed against the mother during the interview.

8. Disagreements and Aggressions - Total—The sum of 6 and 7 above.

[3] For further information see Farina (1958).

9. Failure to Agree—The number of situations for which the parents failed to arrive at a mutually acceptable solution.

10. Verbal Activity—The total amount of time during which the mother and father talked.

While most of the indices comprising the conflict variable appear to have considerable face validity, it may not be apparent why the index Verbal Activity was also used as a measure of conflict. Since the instructional set given to the parents in the interaction interview stressed that they arrive at a joint solution, *S*s continued to speak as long as they remained in disagreement. To test the assumption that Verbal Activity could serve to reflect the difficulty experienced by the parents in reaching agreement or the extent of the disagreement between them, total time talked by both parents in situations where agreement was reached was compared to total time talked in those situations in which they failed to achieve a common solution. Since the difference was in the expected direction and highly significant ($t = 4.76$, $df = 22$, $p < .001$) it may be concluded that *S*s do tend to talk longer in situations where they do not agree.

Reliability of Indices. Four cases were chosen for independent rating by an advanced graduate student in psychology[4] and two of these four cases were rerated by *E* without reference to his previous ratings. The resulting inter- and intrarater correlation coefficients were very similar. For four indices, Speaks First, Speaks Last, Failure to Agree, and Passive Acceptance of Solution, there was virtually perfect agreement. For these, one disagreement occurred in 360 possibilities. For the balance of the indices, the average Spearman *rho* value for inter- and intrarater reliabilities combined was +.89 with a range from +.74 to +.96. Although there are variations from index to index, each of these indices appeared sufficiently reliable to warrant its retention.

RESULTS

The main purpose of this study was to test the hypothesis that there is an association between the sex of the dominant parent and the premorbid adjustment of the schizophrenic son. The analysis of the relevant data was done by deriving indices from the mother and father scores in such a way that negative scores indicated mother dominance and positive scores father dominance. For example, for Speaks First, the number of times the mother spoke first in 12 problem situations was subtracted from the number of times the father spoke first. Table 2 presents the means and standard deviations for each group on each index. Increasingly high positive numbers indicate increasing degrees of father dominance, whereas mother dominance is similarly represented by negative numbers. In the case of Total Time Spoken the mother's score was divided into the father's score so that an index greater than unity indicates father dominance and one less than unity indicates mother dominance. In brief, Table 2 shows that mother dominance was most marked for the Poor group, that the controls were intermediate, and that the Good group was characterized by father dominance.

TABLE 2

MEANS AND STANDARD DEVIATIONS OF GROUPS ON DOMINANCE INDICES[a]

	Controls	Goods	Poors
Speaks First			
Mean	−.5	+5.7	−2.2
SD	9.1	6.3	7.0
Speaks Last			
Mean	−1.0	+3.4	−2.7
SD	5.0	4.0	5.7
Total First and Last			
Mean	−1.5	+9.2	−4.8
SD	10.6	7.1	9.7
Passive Acceptance of Solution			
Mean	−.5	+3.1	−1.2
SD	4.4	3.2	1.9
Total Time Spoken[b]			
Mean	.72	2.0	.72
SD	1.9	3.2	1.3
Yielding-Maximum			
Mean	+.12	+1.3	−.8
SD	2.4	1.1	1.7
Yielding-Minimum			
Mean	−.13	+.8	−1.0
SD	1.9	1.6	1.6

[a] + indicates father dominance; − indicates mother dominance.
[b] A score greater than 1 indicates father dominance; a score below 1 indicates mother dominance.

The significance of the difference among these three sets of means was evaluated by a nonparametric test suggested by Jonckheere (1954). The p values yielded by this test are shown in the first column of Table 3. All differences are significant at at least the .06 level of probability. A further analysis to determine the probability that any two groups were selected from the same population was made by means of the Mann-Whitney *U* test. The p values yielded by this series of tests are presented in the last three columns of Table 3. On this table, p values greater than .05 are listed as nonsignificant (*NS*). For all indices, the Good and Poor groups differed from each other at better than the .01 level of significance in the expected direction. All of the indices

[4] Appreciation is expressed to Richard Dunham for this work.

TABLE 3

LEVELS OF SIGNIFICANCE OF DIFFERENCE AMONG AND BETWEEN GROUPS FOR DOMINANCE INDICES

Index	Among Groups	Between Groups		
		Controls vs. Goods	Controls vs. Poors	Goods vs. Poors
Speaks First	.06	*NS*	*NS*	.01
Speaks Last	.019	.025	*NS*	.01
Total First & Last	.004	.025	*NS*	.001
Passive Acceptance of Solution	.005	.025	*NS*	.001
Total Time Spoken	.0002	.01	*NS*	.001
Yielding-Maximum	.005	*NS*	*NS*	.001
Yielding-Minimum	.004	*NS*	*NS*	.01

in the Good-Control group comparison reveal that the Good group was characterized by more father dominance, and four of the seven indices employed reached a statistically significant difference. Six of the seven dominance indices indicate that the Poor group was more mother-dominated than the Control group, while the seventh index was identical in the two groups. In this last comparison, none of the indices reach the .05 level of significance, however.

The first three columns of Table 4 present the means and standard deviations obtained by the three groups on the ten conflict indices. For nine of these ten indices, extent of con-

TABLE 4

MEANS, STANDARD DEVIATIONS, AND LEVELS OF SIGNIFICANCE FOR GROUPS ON CONFLICT INDICES

	Means and *SD*s			Levels of Significance			
	Control	Good	Poor	Control vs. Total Schizophrenia	Control vs. Good	Control vs. Poor	Good vs. Poor
Frequency of Simultaneous Speech							
Mean	5.4	7.5	11.4	*NS*	*NS*	*NS*	*NS*
SD	4.3	7.5	10.2				
Duration of Simultaneous Speech							
Mean	4.8	9.0	13.8	*NS*	*NS*	*NS*	*NS*
SD	3.6	12.2	16.5				
Interruptions—Mother							
Mean	1.3	3.0	5.6	**	*NS*	**	*NS*
SD	1.5	3.2	4.1				
Interruptions—Father							
Mean	1.2	2.8	4.7	*	*NS*	*	*NS*
SD	1.9	3.3	5.3				
Interruptions—Total							
Mean	2.6	5.8	10.3	**	*NS*	**	*
SD	2.4	6.3	9.1				
Disagreements and Aggressions—Mother							
Mean	1.7	1.5	4.2	*NS*	*NS*	*NS*	*
SD	1.5	2.3	5.1				
Disagreements and Aggressions—Father							
Mean	.8	1.2	2.4	*NS*	*NS*	*	*NS*
SD	1.0	1.6	2.6				
Disagreements and Aggressions—Total							
Mean	2.5	2.8	6.7	*NS*	*NS*	*	*
SD	2.3	3.3	8.0				
Failure to Agree							
Mean	.9	1.0	1.9	*NS*	*NS*	*	*NS*
SD	1.3	1.2	1.4				
Verbal Activity							
Mean	337	424	556	*	*NS*	*	*NS*
SD	106	236	295				

* $p < .05$.
** $p < .01$.

flict was least in Controls, greatest in Poors, and intermediate in the Goods. Only for the 10th index, Disagreements and Aggressions—Mother, did the Control group obtain a somewhat higher score (indicative of greater conflict) than the Goods, although for this index, too, the Poors' score was the highest. In the analysis of these data, each group was compared with every other group, and the Total Schizophrenic group, the Goods and Poors combined, was compared to the Control group. The Mann-Whitney *U* test was employed for these analyses. The resulting *p* levels are indicated in the last four columns of Table 4. The tests employed are one-tailed with the exception of the Good-Poor comparison, where the tests are two-tailed because no prediction of differences in conflict behavior had been made.

The comparisons most relevant to the hypothesis that parents of schizophrenic patients experience more conflict in their interactions than controls are those between the Control and Total Schizophrenic groups. Since all 10 indices were in the expected direction, and 4 of these reached a *p* level of less than .05, the hypothesis appears to be supported. The other comparisons reveal that the Poor group parents experienced more conflict in their interaction than either the Control or Good groups. All 10 indices were greater for the Poors than for the Controls, and 7 reached a statistically significant level. Similarly, all 10 indices were higher for the Poor than for the Good groups, and three reached statistical significance. The Good-Control groups comparisons failed to reveal any statistically significant differences for any index, although 9 of the 10 were higher for the Goods than for the Controls.

A test was made of the hypothesis that the interactions of parents of schizophrenic patients are characterized by dominance of one relative to the other. The dominance indices for these analyses were derived from the mother and father scores so that increasing magnitude of the score indicated an increasingly unequal dominance relationship between the parents. For example, for Speaks First, the submissive member's score, regardless of sex, was always subtracted from the dominant member's score. The index scores were very similar for the two groups and there was no observable directional trend. Four of the seven indices were higher for the Control group, and three were higher for the Total Schizophrenic group.

The PARI scores were very similar for all groups. Neither for mothers nor for fathers could any directional trends be observed on any of the nine scales administered.

DISCUSSION

The results of this investigation revealed a marked relationship between sex of the dominant parent and level of premorbid adjustment achieved by the schizophrenic son. The interactions of parents of good premorbid patients were characterized by father dominance whereas the mother was dominant in the case of parents of poor premorbid patients. However, this correlation does not permit one to assume that such parental behaviors were the cause of the offspring's behavior since this study shares the problems inherent in any retrospective study.[5] Clearly, one can posit alternative explanations for the relations between parental and patient behaviors. For example, it is conceivable that the son's role in the family in the past may have given rise to current parental interactions patterns or that parental awareness of the son's status and their own relationships to the hospital may have served to modify their situational test behavior. Nevertheless, the traits displayed in a situational test such as the one employed in this study may have considerable stability over time and therefore may be representative of earlier behaviors in the family setting. For example, Naka (1948) has demonstrated that certain personality traits are persistent in most individuals over a 15-year time span. Her *S*s were children of 2 and 17 years of age, however.

If it is assumed that the parental behaviors as measured in this study were characteristic of the parents from the time the patient was an infant, how might these have been significant in shaping the offspring's personality? The good premorbid schizophrenic patients displayed more customary adult male behaviors

[5] These and related problems of retrospective studies will be discussed at length in a forthcoming publication (Garmezy, Farina, & Rodnick, 1960).

prior to psychosis than was true of the poor premorbid patients. The family environment in which the good premorbid schizophrenic develops may make it easier for him to achieve a somewhat higher level of maturity and to manifest more adequate male behavioral patterns relative to the poor premorbid schizophrenic patient. Although the good premorbid may still be exposed to such pathogenic family influences as are generated by a tyrannical father and a weak, vacillating mother, the father does appear to assume dominance in the family. Consequently, she is in a position to reward the son for imitating his behavior and to serve as a model, even if a faulty one, for male identification. The poor premorbid schizophrenic, on the other hand, appears to have grown up in a family where the mother exercises control and where the father does not serve as a reasonably adequate male figure with whom the patient can identify. Even in instances where the son makes tentative efforts to identify with the father, the antagonism between the parents may inhibit an effective identification as Mowrer (1950, p. 596) has pointed out. These findings appear to have implications for theories of identification and sex typing, since the behavioral differences between the good and poor premorbid schizophrenic patients are relevant to these concepts.

The findings concerning conflict appear to support the reports by many writers that relationships between parents of schizophrenic patients are characterized by hostility and contentiousness. If their behavior in the situational test can be used to infer their more general behavior, these parents do not appear to be able to cooperate with one another. Impatience with each other and a mutual lack of respect and regard are indicated by frequent interruptions and unwillingness to compromise in the task. This discord may involve the schizophrenic offspring according to Lidz, Cornelison, Fleck, and Terry (1957b), who report that he is often the focus and battleground of parental conflict. Since the stimulus situations employed in this study dealt with problems of disobedience and misbehaviors of children, they may have served as effective cues for eliciting the conflictful parental interactions.

Excessive dominance by one parent, particularly dominance by the wife over the husband, has often been described as the modal parental pattern in families of schizophrenic patients. The results of the present investigation, however, suggest that the premorbid adjustment of the schizophrenic son should be considered in describing the parental dominance relationship. In fact, when the sex of the parent and the premorbid adjustment of the son are disregarded and the Goods and Poors combined are compared to the Control group, no differences in extent of dominance of one parent over the other can be discerned in these samples. Furthermore, since the role the parents play in the child's acquisition of sex-appropriate behaviors appears to vary with the sex of the child (Sears, Pintler, & Sears, 1946), and since the process of identification may be different for females, the sex of the schizophrenic offspring should also be considered in studying the role of possible family influences in the development of a schizophrenic disorder.

SUMMARY

The major purpose of the present study was to investigate a reported association between sex of the dominant parent and the adequacy of premorbid adjustment of the schizophrenic son. The investigation was carried out by means of a situational test so designed and structured that indices of dominance could be obtained. Such indices were length of time spoken, relinquishing one's position in favor of the spouse, etc. Paternal dominance was found to be associated with good premorbid adjustment of the schizophrenic son whereas in the case of those patients whose premorbid adjustment was poor, the mothers were dominant. Because of these differences in parental behaviors, it may have been somewhat easier for the good as compared to the poor premorbid son to learn the more usual adult male behaviors and to achieve a somewhat higher level of maturity.

Extent of conflict present in the parental interactions was inferred from such measures as number of interruptions, number of disagreements, etc. The parents of the schizophrenic patients displayed more conflict than the Control parents, and this conflict was

particularly marked in the parents of the poor premorbid schizophrenic sons. The results of the present investigation suggest that the premorbid adjustment and sex of the schizophrenic offspring need to be considered when investigations of the familial role in schizophrenia are undertaken.

REFERENCES

ARIETI, S. *Interpretation of schizophrenia.* New York: Robert Brunner, 1955.

CERVIN, V. Relationship of ascendant-submissive behavior in dyadic groups of human subjects to their emotional responsiveness. *J. abnorm. soc. psychol.*, 1957, **54**, 241–249.

ELLISON, E. A., & HAMILTON, D. The hospital treatment of dementia praecox. Part II. *Amer. J. Psychiat.*, 1949, **106**, 454–461.

FARINA, A. Patterns of role dominance and conflict in the interaction of parents of schizophrenic patients. Unpublished doctoral dissertation, Duke Univer., 1958.

FRAZER, H. E. Children who later became schizophrenic. *Smith Coll. Stud. soc. Wk.*, 1953, **23**, 125–149.

FRIEDLANDER, D. Personality development of 27 children who later became psychotic. *J. abnorm. soc. Psychol.*, 1945, **40**, 330–335.

FROMM-REICHMAN, FRIEDA. Notes on the development of treatment of schizophrenia by psychoanalytic psychotherapy. *Psychiatry*, 1948, **11**, 263–273.

GARMEZY, N., FARINA, A., & RODNICK, E. H. The structured situational test: A method of studying family interaction in schizophrenia. *Amer. J. Orthopsychiat.*, in press.

GERARD, D. L., & SIEGEL, J. The family background in schizophrenia. *Psychiat. Quart.*, 1950, **24**, 45–73.

JACKSON, P. W. Verbal solutions to parent-child problems and reports of experience with punishment. Unpublished doctoral dissertation, Columbia Univer., 1955.

JACKSON, P. W. Verbal solutions to parent-child problems. *Child Developm.*, 1956, **27**, 339–349.

JONCKHEERE, A. R. A distribution-free K-sample test against ordered alternatives. *Biometrika*, 1954, **41**, 133–145.

KOHN, M. L., & CLAUSEN, J. A. Parental authority behavior and schizophrenia. *Amer. J. Orthopsychiat.*, 1956, **26**, 297–313.

LIDZ, T., CORNELISON, A. R., FLECK, S., & TERRY, DOROTHY. The interfamilial environment of the schizophrenic patient. *Psychiat.*, 1957, **20**, 329–342. (a)

LIDZ, T., CORNELISON, A. R., FLECK, S., & TERRY, DOROTHY. The interfamilial environment of schizophrenic patients: II. Marital schism and marital skew. *Amer. J. Psychiat.*, 1957, **114**, 241–248. (b)

MANN, H. B., & WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.

MATARAZZO, J. D., SASLOW, G., & GUZE, S. B. Stability of interaction patterns during interviews: A replication. *J. consult. Psychol.*, 1956, **20**, 267–274.

MOWRER, O. H. *Learning theory and personality dynamics.* New York: Ronald , 1950.

NAKA, PATRICIA N. Shirley's babies after fifteen years: A personality study. *J. genet. Psychol.*, 1948, **73**, 175–186.

OPLER, M. K. Schizophrenia and culture. *Scient. Amer.*, 1957, **197**, 103–110.

PHILLIPS, L. Case history data and prognosis in schizophrenia. *J. nerv. ment. Dis.*, 1953, **117**, 515–525.

REICHARD, SUZANNE, & TILLMAN, C. Patterns of parent-child relationships in schizophrenia. *Psychiatry*, 1950, **13**, 247–257.

RODNICK, E. H., & GARMEZY, N. An experimental approach to the study of motivation in schizophrenia. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln: Univer. Nebraska Press, 1957. Pp. 109–183.

ROSEN, J. N. *Direct analysis*, New York: Grune & Stratton, 1953.

SCHAEFER, E. S., & BELL, R. Q. Parental attitude research instrument. (PARI). Unpublished manuscript. Lab. Psychol., Natl. Inst. Ment. Hlth., 1955.

SEARS, R. R., PINTLER, M. H., & SEARS, P. S., Effect of father separation on preschool children's doll play aggression. *Child Developm.*, 1946, **17**, 219–243.

STRODTBECK, F. L. Husband-wife interaction over revealed differences. In A. P. Hare, E. F. Borgatta, & R. F. Bales (Eds.), *Small Groups.* New York: Knopf, 1955, 464–472.

TIETZE, TRUDE. A study of mothers of schizophrenic patients. *Psychiatry*, 1949, **12**, 55–65.

(Received December 8, 1958)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 39-46

THE VALIDITY OF HYPNOTICALLY INDUCED VISUAL HALLUCINATIONS[1]

HAROLD WAYNE UNDERWOOD[2]
University of Illinois

STUDIES on the influence of hypnotic suggestion on memory, physical strength, emotional and physiological responses, etc., have established that many hypnotic phenomena involve more than simply acting. One kind of problem that has been particularly formidable, however, has been that of determining to what extent hypnosis can influence perception. Here, the problems involved in studying perception, such as difficulty in insuring that the overt response is an accurate reflection of a percept, are added to problems in studying hypnosis.

The purpose of this study was to test the validity of hypnotic positive[3] and negative[4] visual hallucinations. Operationally, the problem is to determine whether or not there is a percept resulting from a hypnotically suggested visual stimulus, which is similar to the percept resulting from the objective stimulus, and whether "suggesting away" a stimulus produces changes in a percept similar to those produced by objective removal of the stimulus. The primary difficulty in such an investigation is that of insuring that the hypnotic subject (*S*) is not merely faking or pretending to see a suggested stimulus.

METHOD

To test the validity of hypnotic hallucinations, three optical illusions were used in which a field distorts a superimposed figure. For example, in one illusion, a series of inverted Vs (field) superimposed on a square (figure), make the sides of the square appear to widen out toward the top (see Fig. 5). If hypnotic hallucinations are valid, then the positive hallucination of a field on a real figure should cause appropriate figure distortion and likewise, the negative hallucination of an actual field should remove the distortion of the figure. The distortion of figures was measured in terms of quantity as well as simply presence or absence, thereby providing a measure of the degree of effectiveness of the hypnotic suggestion in contrast to objective stimulation. Unfamiliar illusions were used, which were designed so that distortions could not be inferred without perceiving them, unless the *S* had a fairly precise knowledge of the illusions.

Subjects

Hypnotic *S*s were selected by screening approximately 195 *S*s (mostly graduate students) for hypnotic susceptibility by attempting to hypnotize them. Group hypnosis was used on about 100 of those *S*s, in five different groups, and the remaining 95 were seen in individual sessions. *S*s who showed susceptibility in either a group or individual session were seen privately for several sessions to determine whether or not they would be capable of achieving sufficient trance depth to hallucinate. Twelve *S*s were chosen as being suitable for the study on the basis of seeming to have the most vivid hallucinations, while lacking in any gross visual defects or any background that might have made them familiar with the illusions. They all responded positively to trance depth tests of eyelid and arm catalepsies, arm levitation, hand anesthesia, selective amnesia, positive and negative tactual, auditory and visual hallucinations, and posthypnotic suggestions, including trance induction by signal. Also, they were all somnambulists; that is, they could walk, talk, and have their eyes open while remaining in the trance.

Even within this highly select group, six *S*s showed imperfections in their trance performance. Three reported hallucinations that were sometimes "ghost-like" or fluctuating in vividness, and they also reported that hallucinating required considerable effort. Three other *S*s reported very vivid hallucinations, but were obviously not hallucinating with ease. There was much squinting, tear secretion, fatigue and other signs of excessive effort. The remaining six *S*s reported constantly vivid hallucinations (indistinguishable from reality) which were apparently produced with ease.

Since past research has suggested that extremely deep trances are necessary for this type of research (Erickson 1938a) those six *S*s who showed trance imperfections were delegated to a separate group called

[1] This paper is based on a PhD dissertation at the University of Illinois. The author wishes to express gratitude to William M. Gilbert for his generous advice and guidance as chairman of the doctoral committee.

[2] Now at the VAH, Danville, Illinois.

[3] Positive hallucination is the perception of a suggested stimulus, which is in fact not present.

[4] Negative hallucination is the failure to perceive an actual stimulus which is suggested to be absent.

the "deep group." The six who showed flawless performance were called the "very deep group."[5]

The control group of six *Ss* was picked to match the hypnotic groups roughly in age and educational background. No member was to have ever been hypnotized. It was composed of two female and four male *Ss* with a median age of 26.5: three graduate students, one undergraduate, one with a master's degree, and one with a high school education.

The deeply hypnotized group had one male and five female *Ss* with a median age of 27.5. Two *Ss* had bachelor's degrees, one *S* was an undergraduate, one a business college student, and two had a high school education. The very deeply hypnotized group consisted of three male and three female *Ss* with a median age of 22.5: three graduate students, one undergraduate, a nurse, and one with business college training. None of the graduate students was in the field of psychology, and no *S* had any training regarding illusions beyond that obtained in an introductory psychology course.

Apparatus

The illusions were presented to the *Ss* by projecting the figures from behind onto 8″ x 10″ translucent glass screens on which the fields were drawn. A 35 mm. slide projector and frame holders for the screens were mounted on a table with a shield surrounding the screen to block the experimenter (*E*) from the view of *S*.

Illusion A (Osgood, 1953), illustrated in Figs. 1–3, uses a "spoked wheel" field with a square as a figure. The distortion of the square depends on its size and position within the field. When off center of the field to the left (Fig. 1), it is commonly seen as bulging on the right, when directly over center (as in Fig. 2), it has no distortion, and when small and totally to the left of center (Fig. 3), it commonly appears concave on the right and narrowing to the left.

Illusion B (Osgood, 1953), shown in Fig. 4, employs a field of concentric circles which commonly causes a "square" (with slightly convex sides) to appear to have slightly concave sides.

Illusion C (Berliner, A. & Berliner, S., 1948; Henriksson, 1955; Hofmann & Bielschawsky, 1909), which has a field of inverted Vs, is illustrated in Fig. 5 with a square as a figure. In addition to the square 11 trapezoids were used as figures for this illusion (shown in Fig. 6). The effect of the field on each figure is to make the sides seem to angle out to a wider top than is objectively accurate. Even though the field had the general effect of widening each figure toward the top, there were differences among the figures in amount of widening because of the slightly different angle at which the field lines crossed each of the sides of the figures. Since *Ss* could hardly infer this differential illusory effect among the figures, it seemed promising as a control for faking.

Illusion A was used to test the positive hallucination (+H), Illusion B to test the negative hallucination (−H), and Illusion C to test both the +H and −H. A special problem existed in testing the +H in that it was necessary to insure that the *S* was hallucinating the field in the proper position over the figure. To accomplish this, minimal parts of the field were presented on the screen, and the *S* merely had to hallucinate the missing portions. These minimal parts of the field have been called "reference patterns." The reference patterns for Illusions A and C are illustrated in Figs. 7 and 8 respectively, with example figures placed properly.

Ss reported perceptions by a matching technique. A set of figures on a card was placed below the screen on which illusions were presented, and *S* was required to select from that set a figure matching the one observed on the screen. One set of 18 figures was used for Illusions A and B (Fig. 9). From this set there was only one correct matching figure for each illusory figure, and hence responses could be scored as correct or incorrect. A second set for Illusion C (illustrated in Fig. 6) consisted of duplicates of the illusion figures. It may be seen that they formed a progression of figures. The amount of distortion could be reported rather than simply the type.

Conditions

All *Ss* reported perceptions of figures under the five following conditions:

1. Positive hallucination (+H)—figures were shown on the reference pattern with the suggestion that the pattern was filled in to be exactly like the field which he could see on a second screen beside the reference pattern.
2. Negative hallucinations (−H)—figures were shown superimposed on the field with the suggestion that the field was absent.
3. Reference pattern—figures were shown on the reference pattern with no hallucination suggestion. This control was necessary to insure that any distortions of figures in the positive hallucination condition were not caused by the reference pattern alone rather than the hallucinated part of the field.
4. Objective—figures were shown on a blank screen.
5. Illusion—figures were shown on the field. The objective and illusion conditions formed a basis for evaluating the hallucination conditions. If hallucinations are valid, the responses to the +H condition should differ from the objective while approaching or being like the illusion condition, and the opposite should be true for the −H.

In determining the order in which illusions and conditions should be presented, the primary concern was to avoid giving *S* any information at one time that could be used in faking responses later. Obviously, for example, it was necessary to have the +H condition precede the illusion condition.

The order of presentation of the conditions and Illusions A, B and C was: A-reference pattern, A-(+H), B-illusion, B-(−H), C-objective, C-reference pattern, C-(+H), C-(−H), C-illusion, A-illusion, and B-objective with the exception that C-objective and C-reference pattern were reversed with alternate *Ss* in all groups.

[5] The very deep group was selected first and run through the experiment. When the quest for further *Ss* failed to produce *Ss* with flawless trance performance, the deep group was formed and run through the experiment.

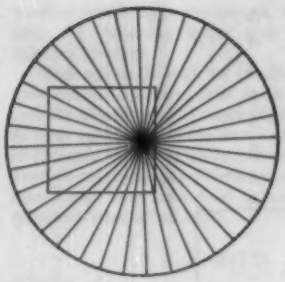

FIG. 1. Illusion A, Illusion Fig. 1.

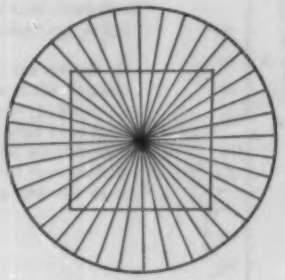

FIG. 2. Illusion A, Illusion Fig. 2.

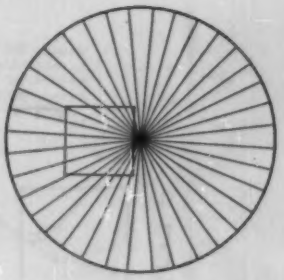

FIG. 3. Illusion A, Illusion Fig. 3.

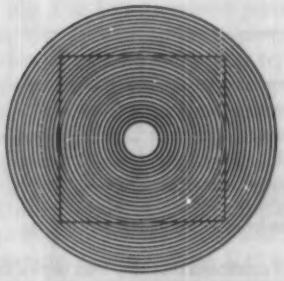

FIG. 4. Illusion B.

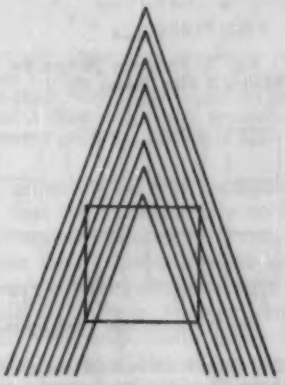

FIG. 5. Illusion C, Illusion Fig. 7.

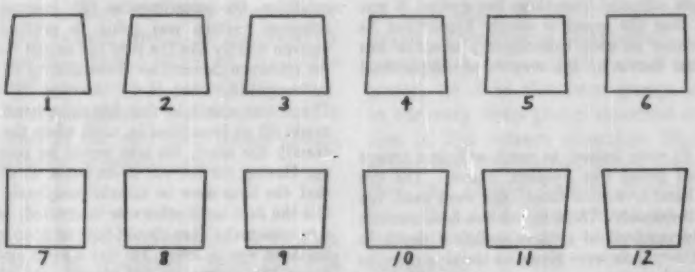

FIG. 6. Figures used with Illusion C

The three figures of Illusion A were always presented in order as numbered. With Illusion C, all the figures (1 through 12) were presented under objective and reference pattern conditions, Fig. 3 through 8 were presented for the hallucination conditions, and Fig. 1 through 8 for the illusion condition. They were presented in different random order for each condition and the same order for all *Ss*.[6]

[6] Figures 9 through 12 were left out when distortion was anticipated since there would not be a matching figure to report. Figures 1 and 2 were left out of the hallucination conditions since the direction of the illusory distortion might be inferred from leaving figures off from only the upper end of the scale.

Hypotheses

On the basis of past research (Benussi, 1927; Erickson & Erickson, 1938b) and a pilot study, it was hypothesized that in very deeply hypnotized *Ss*, the positive hallucination (+H) of the field of an illusion would result in reports of appropriate distortion in the superimposed figure and the negative hallucination

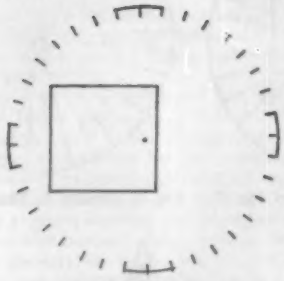

FIG. 7. Reference pattern for Illusion A with Illusion Fig. 2.

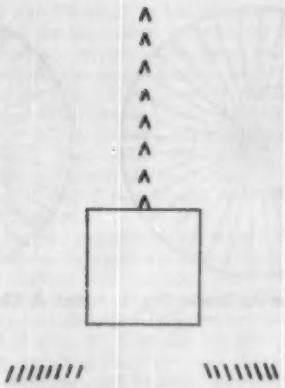

FIG. 8. Reference pattern for Illusion C with Illusion Fig. 7.

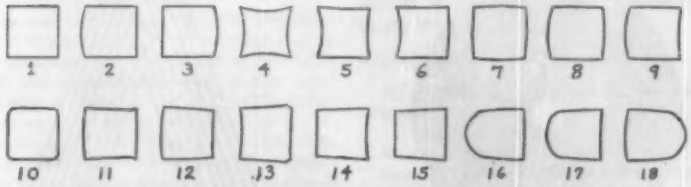

FIG. 9. Figures from which S was to select matches for figures of Illusions A and B.

(−H) of the field would result in reports of a reduction or removal of the figure distortion. Control *Ss* were expected to show less or no such hallucinatory effects. After data were collected from those two groups, it was hypothesized that the group of deeply hypnotized *Ss* would show either no such hallucinatory effect or less effect than that shown by the very deeply hypnotized group.

Procedure

Hypnotic *Ss* were trained to reach sufficient trance depth for their group (see *Subjects*, above). The eye fixation and hand levitation techniques were used, the former more commonly. Three to five one-hour sessions were typically required to achieve sufficient depth in suitable *Ss*. Control *Ss* were given no training prior to the experimental session.

At the experimental session *S* was seated in a chair about three feet from the screen. Hypnotic *Ss* were hypnotized by the technique suitable to them and trance deepening suggestions were given for 5 to 10 min. depending on the particular *S*. They were tested for reports of −H and +H with stimuli unrelated to the illusions and then went through all the conditions while hypnotized. The control *Ss* simply went through all conditions in the nonhypnotized state.

Instructions for all *Ss* on objective, reference pattern and illusion conditions were exactly the same. *Ss* were simply asked to match the projected figures with those below the screen. It was specified with Illusions A and B in all conditions that they were to pick figures showing the kind of shape, though size and amount of curve or angle might be different. With the hypnotic *Ss* in +H condition the experimenter (*E*) instructed that the reference pattern was going to gradually fill in to become exactly like the field (an actual field was beside the reference pattern) as *E* counted to 10 and it would be completely filled in by the time 10 was reached. The *S* was also told that his right hand and forearm would lift as lines filled in, until when the designs were exactly the same, his arm would be pointing straight up. Having carried out those steps, *E* then instructed that the lines were to remain completely filled in just like the field until otherwise instructed; however, if for any reason the lines should fade or change in any way, his hand was to lower. He was also to make judgments about the figures on the screen only when the lines were completely filled in. The figures were then shown on the hallucinated field, and then the hallucination was removed. Similar instructions with appropriate modification were used for −H with the exception that *Ss* were told not only that the screen would become blank, but that they would forget what had been on it before.

Instructions to control *Ss* on the −H and +H conditions were the same as for hypnotic *Ss*, except that they were told to *imagine* the lines being filled in or absent and they were to *guess* the forms the projected figures would appear to have, if the suggested stimulus condition actually existed. Throughout the instructions, there were minor differences from the

wording for the hypnotic groups, which were consistent with *S*'s imagining rather than *E* asserting actual stimulus changes.

After going through all conditions, the control *Ss* were questioned about prior knowledge of the illusions and the effectiveness of imagining the stimuli. Hypnotic *Ss* were questioned in the trance about prior knowledge of illusions, clarity of hallucinations, and whether they ever reported percepts different from what they actually saw. They were instructed to awaken and the same questions were asked again.

RESULTS

Illusion A

In the +H condition (Illusion A was not used to test the negative hallucination), the only *Ss* reporting correct distortions were in the very deep trance group. In that group two of the six *Ss* reported correct distortions on illusion Fig. 1 and also reported perceiving concaveness, though missing the narrowing on Fig. 3. A third *S* in that group also reported the same perceptions, but then changed his mind stating that all three figures looked like squares. Regarding Fig. 2 in which there is no noticeable illusory distortion, the number of *Ss* reporting the correct percept (square) in the control, deep trance, and very deep trance groups were respectively three, six, and four.

All *Ss* reported seeing squares (i.e. no distortion) in the reference pattern condition and the illusory effect in the illusion condition.

Statistically, neither the very deep trance nor the deep trance group is significantly different from the control group on any of the conditions (Fisher's exact probability test). Thus, the results suggest a tendency toward a +H effect only in the very deep trance group, which is statistically insignificant.

Illusion B

No *S* gave the correct response to Illusion B under the −H condition. Since the +H condition was not used with this illusion, evidence here pertains only to −H, and is negative. Reported perceptions on the objective condition were veridical and the illusory effect was reported on the illusion condition.

Illusion C

Results from Illusion C are presented in Fig. 10 and Table 1.

Regarding Fig. 10, it should be remembered that plus error is normal illusory error. First

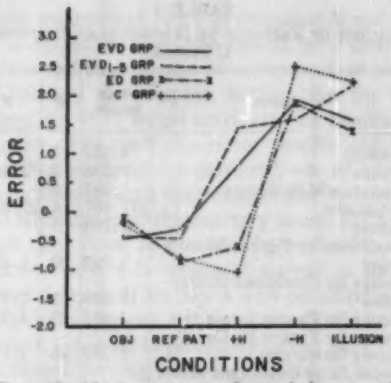

FIG. 10. Mean error per figure judgment with Illusion C, Fig. 3-8, in the very deeply hypnotized group (Evd), very deeply hypnotized group excluding deviant *S* (Evd 1-5), deeply hypnotized group (Ed) and control group (C). Error is in figure units.

with respect to the −H condition, it may be seen that there is essentially no evidence for negative hallucinations (−H responses do not deviate from illusion toward objective responses), a finding consistent with results from Illusion B. There is, however, clearly a trend in the very deep trance group indicating a tendency toward the effects expected with positive hallucinations (+H responses deviate from objective toward illusion responses), and no such effect in the control or deep trance group. Considering +H effect in terms of number of *Ss* showing it, one *S* in the control group, no *S* in the deep group, and four *Ss* in the very deep group reported figure distortion in the correct direction. Figure 10 also shows that the control group tended to guess that the illusion caused distortion opposite from that actually produced, as indicated by the error scores under both the +H and −H conditions.

In order to evaluate these and other trends, a three dimensional analysis of variance was performed. As indicated in Table 1 with all 18 *Ss*, the only significant *F* value results from variance introduced by conditions. The index to the hallucinatory effect is the group by condition interaction, and the obtained *F* ratio would be expected by chance slightly more than 10% of the time. The responses of one *S* from the very deep group are sufficiently deviant from those of all other *Ss* that a con-

TABLE 1
ANALYSIS OF VARIANCE OF JUDGMENTS OF ILLUSION C FIGURES

Source	Mean Square	df	F
Subjects	4.482	17	
Groups	6.236	2	1.468
Subjects within Groups	4.247	15	
Conditions	177.467	4	28.522*
Figures	1.193	5	
Conditions by Figures Interaction	1.999	20	1.126
Groups by Conditions Interaction	10.638	8	1.710
Groups by Figures Interaction	2.415	10	1.730
Groups by Figures by Conditions Interaction	1.257	40	1.116
Pooled *Ss* by Conditions within Groups Interaction	6.222	60	
Pooled *Ss* by Figures within Groups Interaction	1.395	75	
Pooled *Ss* by Conditions by Figures within Groups Interaction	1.126	300	

* Significant at .001 level.

siderably more complete description of the results is gained by presenting analyses of both the complete data and data with that case excluded. This *S* was unique in reporting the opposite from correct perceptions on +H, −H and illusion conditions on all three illusions. His responses to the reference pattern condition on Illusion C were very erratic, some being markedly in error in both directions and others fairly accurate. Considering all 12 figures of Illusion C, he increased the variance of the very deep group on reference condition from 1.5 to 33.9. Poststudy questioning revealed that a few years prior to the study he had been hospitalized for a hysterical reaction. By excluding the deviant *S*, the *F* ratio from group by condition interaction increases to the point that it would be expected by chance less than 5 times in 1000. Also the variance due to groups gains considerable significance, barely missing the .05 level. Thus, there is no conclusive evidence for a group by condition interaction, but a tendency toward such an effect is present, and the tendency is given further weight when the inordinate contribution of one *S* is taken into account.

To further isolate the source of variance in the group by condition interaction, *t* tests were performed, contrasting the groups on the various conditions. Among the group comparisons, the only significant *t* ratios are found on the +H condition. The contrast between the control group and the entire very deep group reaches the .05 level of significance; when the deviant *S* is excluded, the contrast reaches the .005 level (one alternative *t* tests).

When the differences between conditions are tested over all three groups together, each condition is significantly different from each other condition, except for the contrasts of the objective vs. positive hallucination, reference pattern vs. positive hallucination, and negative hallucination vs. illusion. The significance levels by the sign test are greater than .001 except for the objective vs. reference pattern contrast which is .004.

When the differences between conditions are tested within groups, there are three reversals to the results over the three groups together. The objective vs. reference pattern contrast is insignificant by the *t* test in the very deep group, though it remains significant if the deviant *S* is excluded. And objective vs. positive hallucination and reference pattern vs. positive hallucination contrasts are significant in the very deep group excluding the deviant *S* (.05 and .025 respectively by one alternative *t* tests).

One peculiar outcome was that figures were judged to be significantly more narrow toward the top in the reference pattern condition than in the objective condition.

Poststudy Questioning

No subject reported having seen any of the illusions prior to the study and all answers were the same in and out of the trance with the exception that only while hypnotized, one *S* from the very deep group reported possibly having seen Illusion A (none of his +H responses were correct, however, on that illusion). All in the control group reported being unable to imagine lines appearing or disappearing. In the very deep group, all *Ss* reported complete disappearance of lines during the −H. Four of those *Ss* reported complete +Hs and the other two reported a slight fading of lines at times during the +H, in spite of having failed to report such fading at the time. These were the two *Ss* who only showed hallucinatory effects part of the time. In the deep group, three *Ss* reported complete −Hs and +Hs.

and the remaining three *S*s reported blurring of lines beyond recognition during the −H and faded or incomplete +Hs.

There was thus no evidence that *S*s had any usable prior information about the illusions. The reports of positive hallucinations were generally reflected in the objective measures, although there were several cases in which claims of hallucinations were not borne out. Even though *S*s generally reported having negative hallucinations, objective indication of their occurrence was lacking.

DISCUSSION AND CONCLUSIONS

The evidence is quite conclusive that reported negative hallucinations failed to alter perception of the field, to a sufficient extent to lessen measurably its distorting effect on the figure. There is, in fact, no evidence that there was any perceptual alteration at all accompanying the −H. One may not conclude that there was none, however, since perceptual alteration without influence on the illusion is conceivable. Relative to this question, evidence exists which indicates that illusory processes occur both retinally and cortically (Motokawa, 1950). If the hypnotic −H is cortical in locus, then the field processes might have already distorted the figure in the retina at a point prior to the locus of hypnotic influence.

There is conclusive evidence that the field +H did not have the same effect on the *S*s as did objective field stimulation. Even when *S*s were in a very deep trance, their reports that the +H was indistinguishable from the actual field were not borne out by the objective measures of illusory effect. It must be recognized, however, that the study measured only a limited aspect of hypnotic hallucinations.

No definite conclusions can be reached from the study as to whether perception was altered by the +H. *S*s in a trance of a depth often accepted as sufficient for research in deep trance behavior did not show measurable figure distortion accompanying +Hs of the fields, even when they reported that the +H and actual field were indistinguishable.

There is suggestive evidence, however, that the positive hallucination of a field for *S*s who attained very deep trance did cause figure distortion. This trend was manifest in both measures of the +H (illusions A and C); it is magnified when the results for a deviant *S* are excluded.

Perhaps the simplest explanation for the trend for +H would attribute the results to chance. The near significant results and the inordinate effects of a deviant *S* are, of course, arguments against that explanation. A further bit of evidence to the contrary comes from the fact that three of the four hypnotic *S*s who showed any +H tendency showed it on all three figures of Illusion A and on Illusion C. The remaining two *S*s showed no signs whatsoever of the +H; in poststudy questioning they admitted having had incomplete hallucinations. A more random occurrence of correct responses would be expected in results due to chance.

Also supporting the idea that valid +Hs were obtained is the fact that among all the hypnotic *S*s, the four who showed +H effects had complete spontaneous amnesia, while all other *S*s had no such amnesia. The implication is that inadequate trance depth is responsible for the inconclusive results. Spontaneous amnesia, thorough training for a deep trance, and behavior that is completely consistent with a deep trance, would seem to be the minimal criteria for adequate trance depth.

Among the extraneous variables that might have been influential, major consideration should be given to the possibility that the *S*s inferred the correct responses from past knowledge or from cues in the experimental situation. Hypnotic *S*s are reputed to be particularly adept at using such information in faking responses. Although this study was designed specifically to prevent such an occurrence, it is not inconceivable that *S*s under hypnosis could recall some obscure childhood experience with the illusions or some such event that would allow them to fake their responses. This type of explanation does not, however, account for the fact that there was no evidence whatsoever for a valid −H while there was suggestive evidence for the +H. Any past knowledge or cues from the experiment should have been equally applicable under both conditions.

It is of considerable importance that the control and hypnotic groups were not made comparable prior to the initiation of experi-

mental procedures. In selecting hypnotic *S*s, rigid criteria of trance depth were applied and the *S*s used in the study were screened from a very large number of potential *S*s. No such intensive screening could be used in selecting control *S*s. In addition to the fact that it is extremely difficult to find *S*s capable of very deep trance, it was felt that had controls been selected and trained to reach a deep trance as were the hypnotic groups, then one could not be certain that they would not enter a trance, perhaps inadvertently, under conditions of waking suggestion, thereby nullifying their function as a control group. The control group was selected solely for comparability in age and education level. It is quite possible that the groups could have differed in some relevant but uncontrolled way. In particular, the highly selective procedure used in obtaining hypnotic *S*s may have selected *S*s with unusually vivid visual imagery. Such a capacity might make it possible for them to imagine the illusion field with sufficient vividness to perceive the correct figure distortion whether they were hypnotized or not. This conjecture is, of course, contrary to the reports of very deeply hypnotized *S*s that in hyponosis, their hallucinations are indistinguishable from reality and when not hypnotized they only have fleeting mental images.

SUMMARY

The validity of hypnotic visual hallucinations was investigated by the use of three optical illusions, in which a field of lines distorts a superimposed figure. If hallucinations actually involve perceptions similar to those aroused by objective stimulation, then a figure superimposed on the hallucinated field should appear properly distorted and the distortion of a figure in an actual field should be removed if the absence of the field is hallucinated.

Six deeply hypnotized *S*s and six very deeply hypnotized *S*s reported perceptions of figures projected on a screen during both hallucinated and actual presence and absence of the field on the screen. Six nonhypnotized *S*s performed in the same conditions except that they imagined the presence and absence of the field and guessed the way figures would look if the suggested stimulus conditions actually existed.

1 Conclusive evidence was obtained that the negative hallucination did not measurably lessen the distorting effect of the field.

2. Conclusive evidence was obtained that for *S*s who satisfy only minimal criteria for hallucinations and deep trance, the positive hallucination of the field does not cause measurable figure distortion.

3. Suggestive evidence was found that for *S*s who are in a very deep trance, in which all behavior is thoroughly consistent with a deep trance and all visual hallucinations are produced with ease, the positive hallucination of a field does cause figure distortion, but to a degree less than that caused by objective field stimulation.

REFERENCES

BENUSSI, V. Zur experimentelin Grundlegung hypnosuggestiver Methoden psychischer Analyse. *Psychol. Forsch.*, 1927, **9**, 197–274.

BERLINER, A., & BERLINER, S. The distortion of straight and curved lines in geometrical fields. *Amer. J. Psychol.*, 1948, **61**, 153–166.

ERICKSON, M. H. A study of clinical and experimental findings on hypnotic deafness: II. Experimental findings with a conditioned reflex technique. *J. gen. Psychol.*, 1938, **19**, 151–167. (a)

ERICKSON, M. H., & ERICKSON, E. M. The hypnotic induction of hallucinatory color vision followed by pseudo-images. *J. exp. Psychol.*, 1938, **23**, 581–588. (b)

HENRIKSSON, M. The effect on an established percept of a perceptual process beyond awareness. *Acta Psychol.*, 1955, **11**, 346–355.

HOFMANN, B., & BIELSCHAWSKY, A. Über die Einstellund der scheinbaren Horizentalen und Vertikalen bei Betrachtung eines von schrägen Konturen erfülten Gesichtsfelds. *Arch. ges. Physiol.*, 1909, **126**, 453–475.

MOTOKAWA, K. Field of retinal induction and optical illusion. *J. Neurophysiol.*, 1950, **13**, 413–326.

OSGOOD, C. E. *Method and theory in experimental psychology*. New York: Oxford Univer., 1953.

(Received April 29, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 47–51

PERSONALITY SIMILARITY AND FRIENDSHIP[1]

CARROLL E. IZARD[2]
Vanderbilt University

WHY DO Person A and Person B find each other attractive and each other's company a satisfying experience? Two antithetical answers have been formulated to this question in recent years. Winch (1958) argued that the principle of complementarity explains attraction between marriage partners: successful marriages are made up of people whose needs and characteristics complement each other. Newcomb (1956), working with college males who were brought together as strangers and shared common quarters for several months, pointed to perceived similarity as the most significant factor in interpersonal attraction: Person A is attracted to Person B if A perceives B as like A. Newcomb's similarity thesis was given some support by Kelly (1955), who, like Winch, used married couples as *Ss*. Several other studies have found real or assumed similarity to be a significant factor in interpersonal relations (Fiedler, 1958; Rosenfeld & Jackson, 1959); however, Morton (1959)[3] reviewed the literature concerned specifically with personality similarity and friendship and found contradictory conclusions.

The principal thesis underlying the present study is a variant of Newcomb's similarity principle but differs in its emphasis on similarity of affective personality factors and on actual rather than perceived similarity. The affective component in interpersonal attraction has come to seem more important with the recent work of Tagiuri (1958) and Schutz (1958) on the significance of affective rapport—"mutuality," "congruency," "compatibility"—in interpersonal relations and group productivity.

These developments support the basic postulate of the present work, that interpersonal positive affect—the expression of favorable feeling, self-involving interest, acceptance, or esteem in relation to another person—is a significant determinant of individual and interpersonal behavior. A further underlying assumption is that personality similarity facilitates the expression of interpersonal positive affect.

The aim of this study was to throw some light on the general problem of interpersonal attraction by studying personality (affective) variables in relation to friendship. The specific hypotheses were (*a*) mutual friends have similar personality profiles, and (*b*) mutual friends have significant positive correlations on some of the separate personality characteristics that make up the profile.

METHOD

Over 200 students from a high school and a private college, both of which draw largely from the white middle and upper classes, were asked to list their closest personal friends in rank order. They were then given Edwards' (1954) Personal Preference Schedule (PPS), a forced choice personality inventory that measures 15 manifest needs originally defined by Murray (1938). Various PPS scales have been found significantly related to other personality tests (Silverman, 1957) and to relevant criteria derived from actual behavioral situations (Bernardin & Jessor, 1957; Gisvold, 1958; Izard, in press).

The sociometric data made possible the selection of 30 pairs who chose each other as best friends. In addition, a control group of 60 *Ss* from an entering college class were selected and paired at random, with the restriction that there be the same number of male and female pairs as in the 30 pairs of friends.

An analysis of variance technique, described in detail in connection with presenting the results, was used to test the hypothesis that mutual friends have similar personality profiles. This method made it possible to test for profile similarity between pairs of friends and between random pairs, and to compare the group of friend pairs to the group of random pairs with respect to similarity. The technique gives considerably more information and tests of significance based on more degrees of freedom than would the more conventional

[1] This paper was presented at the 1959 meeting of the APA.

[2] The author is indebted to C. I. Bliss, Yale University and Connecticut Agricultural Experiment Station, for invaluable assistance in developing the analysis of variance procedure for comparing profiles; and to Dudley Peeler, Edna Johnson, David Catron, and Billie Farmer for assistance in obtaining and analyzing the data for this study.

[3] Gulliksen was principal investigator under Office of Naval Research (Contract Nonr 1858(15)), and National Science Foundation Grant G-642, Princeton: Princeton University and Educational Testing Service, 1959.

TABLE 1
ANALYSIS OF VARIANCE OF PPS SCORES OF FRIEND AND RANDOM PAIRS

Source	*df*	Sum of Squares	*MS*	*F*	*P*
Between Variables	14	$\Sigma V_i^2/N - T^2/N_o = A$	805.338	47.803	.001
Variables × Groups	14	$(\Sigma v_{i1}^2 + \Sigma v_{i2}^2)/N_g - C - A$	16.847	<1.000	
Variables × Pairs (Friends)	406	$\Sigma y_{i1}^2/N_p - \Sigma v_{i1}^2/N_g$	23.143	1.529	.01
Variables × Pairs (Random)	406	$\Sigma y_{i2}^2/N_p - \Sigma v_{i2}^2/N_g$	22.811	1.226	.05
Variables × *S*s (Friends)	420	$\Sigma y_{i1}^2 - \Sigma y_{i1}^2/N_p$	15.138	1.229	.05
Variables × *S*s (Random)	420	$\Sigma y_{i1}^2 - \Sigma y_{i2}^2/N_p$	18.599		
C (Correction)	1	T^2/N_o			

series of profile correlations or D measures. An r or D is useful as an index of similarity for a single pair, but neither is as powerful as the analysis of variance technique for studying a series of pairs and for comparing one such series with another with respect to profile similarity.

The hypothesis that mutual friends have significant correlations on some of the individual personality variables was tested by computing intraclass correlation (r_i) for each of the 15 personality characteristics. Corresponding correlations were computed for the *S*s paired randomly. The significance of an r_i was determined by Fisher's (1948) method: transform r_i to z, compute corrected z or $z_c = z + \frac{1}{2n' - 1}$, and then compare z_c with $z_{c_o} = \frac{1}{n' - \frac{3}{2}}$. The acceptable level of significance was set at .025, since testing the hypothesis required the computation of 30 r_i's.

RESULTS

The hypothesis that mutual friends have similar personality profiles was supported by the analysis of variance for comparing profiles. For this analysis, profiles were entered in adjacent rows of the data sheet. Since entries in a given row were not independent and summed to a constant, the terms usually based on row totals drop out. To compute the terms in the analysis, let y_1 and y_2 represent the scores of friends and random *S*s respectively on a given personality variable, y_{i1} and y_{i2} the respective totals for a friend and random pair on a variable, v_{i1} and v_{i2} the variable total for the friend group and random group respectively, V_i the total for both groups on a variable, T the grand total for all observations, N_p the number in a unit (pair), N_g the number in a group, N the number of *S*s, and N_o the total number of observations. The analysis of variance together with the computational formulas for the sums of squares is presented in Table 1.

The Between Variables term simply represents the variance associated with the average profile for the 120 *S*s. This term is of no consequence in the present study. It simply indicates that on the average the *S*s responded significantly differently to the 15 PPS scales. It was expected to be quite large in comparison with its error—the Variables × Groups term.

The Variables × Groups term is similar to a simple Between Groups term in a single variable analysis of variance. The principal difference is that here each group is measured on 15 separate variables. Hence, the variance due to variables can be isolated (Between Variables term) and the groups compared in terms of their differential interaction to the 15 variables (Variables × Groups term). The same logic applies to the other interaction terms in the analysis. The Variables × Groups term was tested against the pooled variance of the two Variables × Pairs terms. The resulting F was less than 1.000. This indicated that the average profile for the friend group did not differ significantly from that of the random group.

The test for profile similarity of friends

$$F = \frac{\text{Variables} \times \text{Pairs (Friends)}}{\text{Variables} \times \textit{S}\text{s (Friends)}}$$

can be interpreted as a between variance—within variance ratio. The variance within pairs of friends (Variables + *S*s) was significantly smaller ($P < .01$) than the variance between friends (Variable × Pairs). However, the corresponding F for the random *S*s was also significant, though less so ($P < .05$). It was considered that this significant F for similarity

of random pairs was possibly due to chance fluctuations in random sampling. However, two additional samples of random pairs from the same population had significant profile correlations. While further research is needed to determine the meaning of this finding, it makes it necessary to test whether friend pairs are significantly more alike than *Ss* paired at random.

In order to compare the relative similarity of friend and random pairs, it was necessary to check whether the two groups were homogeneous with respect to overall variability. This test

$$F = \frac{\text{Variables} \times \text{Pairs (Random)} + \text{Variables} \times \text{Ss (Random)}}{\text{Variables} \times \text{Pairs (Friends)} + \text{Variables} \times \text{Ss (Friends)}}$$

($P > .025$), indicated that the profiles of the 60 *Ss* in the friend group did not differ on the average from those of the 60 *Ss* in the random group. The comparison of the within variances of the two groups of pairs

$$F = \frac{\text{Variables} \times \text{Ss (Random)}}{\text{Variables} \times \text{Ss (Friends)}}$$

($P < .05$), showed that mutual friends considering a pair as a unit, were significantly more similar than *Ss* paired at random.

Table 2 presents the intraclass correlations testing the hypothesis that mutual friends have significant positive correlations on some of the personality characteristics that make up the profile. As shown in the table, no r_i was significant for the *Ss* randomly paired. For mutual friends, r_i was significant at the .025 level for Exhibition, at approximately the .01 level for Deference, and beyond the .01 level for Endurance.

The results from both the analysis of variance and intraclass correlation are evidence for similarity and contrary to what would be expected from complementarity; namely, that the scores of friend pairs on a particular variable should correlate negatively. However, they do not completely rule out the possibility of another less well defined type of complementarity, in which a given personality characteristic (e.g., Dominance) complements another (e.g., Abasement). To test for the presence of such complementarity across characteristics, it is necessary to identify characteristics that are psychologically complementary. There is little theoretical or empirical basis for making these pairings. The most appropriate empirical basis available for these pairings is the magnitude of the negative intercorrelations of the 15 variables in the standardization group. These negative correlations might be interpreted as evidence of within person complementarity. Primarily on this basis, the seven most relevant of the 105 possible combinations were tested. The intraclass correlation is again appropriate, since there is no logical basis for assigning *Ss* to x and y columns for an ordinary product moment correlation (Fisher, 1948). The estimate of correlation for a given pair of characteristics would be obtained by pooling the two r_is for that combination, and the acceptable level of P would be somewhat less than .01. The 14 r_is were essentially zero, the largest having a sign in the opposite direction to that predicted by the complementarity principle.

There is a possibility that the social desirability of characteristics may be a factor in the friendship pairings, although the fact that Edwards employed the forced choice technique to eliminate social desirability at the item level makes this unlikely. To check on the possibility, the 98 students in a general psychology course were asked to rank the 15 characteristics in terms of social desirability. The final ranks based on the *Ss*' individual

TABLE 2
INTRACLASS CORRELATIONS FOR FRIENDS AND FOR RANDOM *Ss*

PPS Variable	Intraclass Correlations	
	Friends	Random *Ss*
Achievement	.200	.031
Deference	.435*	.347
Order	.279	−.157
Exhibition	.386*	.179
Autonomy	−.001	.060
Affiliation	.022	.138
Intraception	.113	−.183
Succorance	−.082	−.077
Dominance	.112	.201
Abasement	.168	.296
Nurturance	.100	.223
Change	.253	.212
Endurance	.473*	−.248
Heterosexuality	.116	.193
Aggression	.167	−.049

* $P < .025$.

rankings were normalized by means of Fisher's rankit transformation (Fisher & Yates, 1953). Since the rankits for the 49 men correlated .984 with the corresponding values for the 49 women, final rankits for social desirability were based on all 98 *Ss*. The rankits for social desirability were correlated with rankits derived from the rank order of the characteristics in terms of the magnitude of the r_s for the friend group. This correlation was only .002, indicating that on the average there is no relationship between the social desirability of a given personality characteristic and the degree to which friends are similar on that characteristic. Further, it is interesting to note that the profile based on this index of social desirability did not correlate significantly with the obtained average profile for a large sample of college males or with the average profile for a large sample of college females.

DISCUSSION

There is some support for the importance of perceived similarity, assumed similarity, and actual similarity in interpersonal attraction and friendship. Newcomb (1956) found that perceived similarity was a key factor in interpersonal attraction, and Fiedler (1958) reported that assumed similarity was significantly related to leader attitudes and team effectiveness. Rosenfeld and Jackson (1959) showed that similarity on three scales from the Guilford-Zimmerman Temperament Survey (Sociability–Unsociability, Ascendance–Submissiveness, Security–Insecurity) was significantly related to friendship ratings of female office employees. The present study showed that mutual friends, people who are attracted to one another over an extended period of time, have personality profiles as measured by the 15 scale PPS that are significantly more similar than those of *Ss* paired at random and that they have significant positive correlations on some of the separate personality characteristics.

The question arises as to why personality similarity is a factor in friendship or interpersonal attraction and as to how this factor operates in interpersonal dynamics. The role of personality similarity may be understood by viewing it in relation to the underlying hypotheses of the present study. It was postulated that interpersonal positive affect is an important determinant of interpersonal behavior and that personality similarity facilitates the expression of positive affect. It seems reasonable that the expression of positive affect—favorable feeling, self-involving interest, acceptance, esteem—enhances interpersonal attraction. But how does personality similarity facilitate positive affect? If Person A and Person B have similar personality profiles, they have vested affect or affect needs in common. More dynamically, they have similar ways of expressing and receiving affect, and this actual similarity of affective characteristics might reasonably be expected to facilitate mutually satisfying interactions and experiences. For one thing, the same individuals, groups and interpersonal situations that evoke positive affect in one member of an affectively similar pair would tend to evoke positive affect in the other. Actual personality similarity should also increase the accuracy of interpersonal perception and communication, particularly the perception and communication of attitudes, values and other affective aspects of experience. This is almost tantamount to saying that *Ss* can perceive and communicate that which is similar to their own experience more readily than the dissimilar, the familiar more readily than the unfamiliar. This formulation is in keeping with Newcomb's (1956) conception that accurate interpersonal communication and perception of similarity are interdependent rewarding conditions. Insofar as a person sees his self-involving, affect laden desires and actions confirmed by the desires and actions of another, his perceptions of self and others receives confirmation. This line of thinking is also consonant with the concept of congruency—"the tendency, correctly based on experience, to perceive a person's feelings for us as congruent with our feelings for him"—the factor which Tagiuri (1958, p. 321) has considered the most significant determinant of social preference. In fine, personality similarity, or similar affect needs and similar ways of expressing and receiving affect, should facilitate interpersonal positive affect and make for effective communication, compatibility and cohesiveness in the two person group.

SUMMARY

The present study was designed to test the hypotheses that mutual friends (*a*) have similar personality profiles, and (*b*) significant positive correlations on some of the separate personality characteristics that make up the profile. Mutual "best" friends were determined by a sociometric procedure, and personality characteristics were measured by the 15 scale PPS.

An analysis of variance technique for comparing personality profiles showed that pairs of friends were significantly more similar than pairs established by random assignment. Hypothesis *b* was supported by significant intraclass correlations among friends on Exhibition, Deference, and Endurance and no significant correlations among the random *S*s. These results tend to rule out the possibility of within characteristics complementarity, and an additional analysis yielded evidence against across characteristics complementarity. Further analysis showed that the social desirability of the personality characteristics was not significantly related to the degree to which friends are alike on the characteristics or to the degree to which the characteristics are actually present in the population sampled.

Accepting friendship as a criterion of attraction between people and the PPS as a measure of affective characteristics, the present findings may be restated as follows: personality similarity or similarity of affect needs and of ways of expressing and receiving affect is a significant factor in interpersonal attraction. These findings were discussed in relation to Newcomb's similarity principle and Tagiuri's concept of congruency. Personality similarity was interpreted as essentially a facilitator of interpersonal positive affect, the latter being postulated as a key determinant of attraction as well as other aspects of interpersonal behavior.

REFERENCES

BERNARDIN, A. C., & JESSOR, R. A construct validation of the Edwards Personal Preference Schedule with respect to dependency. *J. consult. Psychol.*, 1957, **21**, 63–67.

EDWARDS, A. L. *Manual, Edwards Personal Preference Schedule.* New York: Psychological Corp., 1954.

FIEDLER, F. E. *Leader attitudes and group effectiveness.* Urbana: Univer. Illinois Press, 1958.

FISHER, R. A. *Statistical methods for research workers.* London: Oliver, Boyd, 1948.

FISHER, R. A., & YATES, F. *Statistical tables for biological, agricultural and medical research.* New York: Hafner, 1953.

GIEVOLD, D. A validity study of the Autonomy and Deference subscales of the EPPS. *J. consult. Psychol.*, 1958, **22**, 445–447.

IZARD, C. E. Personality characteristics associated with resistance to change. *J. consult. Psychol.*, in press.

KELLY, E. L. Consistency of the adult personality. *Amer. Psychologist*, 1955, **10**, 659–681.

MORTON, A. S. Similarity as a determinant of friendship: A multidimensional study. *USN tech. Rep.*, 1959, Nonr 1858 (15), Princeton.

MURRAY, H. A. *Explorations in personality.* A clinical and experimental study of fifty men of college age. New York: Oxford, 1938.

NEWCOMB, T. M. The prediction of interpersonal attraction. *Amer. Psychologist*, 1956, **11**, 575–586.

ROSENFELD, H., & JACKSON, J. Effect of similarity of personalities on interpersonal attraction. *Amer. Psychologist*, 1959, **14**, 366–367 (Abstract).

SCHUTZ, W. C. *FIRO: A three dimensional theory of interpersonal behavior.* New York: Rinehart, 1958.

SILVERMAN, R. E. The Edwards Personal Preference Schedule and social desirability. *J. consult. Psychol.*, 1957, **21**, 402–404.

TAGIURI, R. Social preference and its perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford, Calif.: Stanford Univer. Press, 1958.

WINCH, R. F. *Mate-selection: A study of complementary needs.* New York: Harper, 1958.

(Received June 8, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 52-60

CORTICAL INHIBITION AND EXTRAVERSION-INTROVERSION

WESLEY C. BECKER
University of Illinois[1]

THE RESULTS of several related experiments have led Eysenck to hypothesize that cortical inhibition is one of the bases for individual differences in extraversion-introversion. The extravert is assumed to generate more cortical inhibition faster and to dissipate it more slowly than the introvert. According to Eysenck (1955), cortical inhibition builds up in any neural structure mediating a stimulus-response connection as a result of excitation of that structure. Eysenck has further assumed that Köhler and Wallach's (1944) satiation concept and Hull's (1943) reactive inhibition concept can be partially subsumed under the concept of cortical inhibition. Satiation effects are usually inferred by measuring the effect of stimulation on subsequent perceptual responses. Reactive inhibition is usually inferred by measuring the effect of making a response on subsequent motor responses. Köhler and Wallach considered their satiation effects to be operating within the central nervous system, while Hull referred only to "physical structures" as the locus of action for reactive inhibition, leaving open the question of the effects as being neural or muscular or both.

Eysenck has at times used the terms reactive inhibition, cortical inhibition, and satiation interchangeably, thus neglecting the historical differences in meanings and fostering some confusion. In one paper (1955), cursory reading would lead one to believe that Eysenck has fully embraced Hull's reactive inhibition concept as the primary construct in his theory, but this is not actually the case, since he modified the definition of reactive inhibition to rule out muscular effects. The resulting construct is more appropriately called *reactive cortical inhibition*. This concept, as used by Eysenck, is assumed to cover what is common to the operations involved in defining satiation and reactive inhibition, namely, the effect on subsequent behavior of the repeated excitation of a neural structure mediating a stimulus-response connection. However, this concept is not assumed to cover the additional operations (such as magnitude being dependent on the amount of work involved) which might be used to define reactive inhibition or satiation.

Eysenck has published two studies which have supported the hypothesis that reactive cortical inhibition is one of the bases for individual differences in extraversion-introversion. In the first study (1955), he found significant differences between male hysterics (i.e., extroverted neurotics) and dysthymics (introverted neurotics) in the amount of aftereffect produced by kinesthetic stimulation on tactile judgments of size. The greater cortical inhibition effects produced in extraverts was believed to have led to decreased judgments of size following stimulation. In this study, Guilford's Rhathymia Scale was used in addition to psychiatric diagnosis to define extraversion-introversion. In a second study using normal male university students, Eysenck (1956) argued that extraverts with greater reactive cortical inhibition show more reminiscence on a pursuit rotor task than introverts. In this experiment, the operations defining cortical inhibition paralleled those for reactive inhibition, whereas in the first experiment the operations paralleled those for satiation. His findings indicated a significant correlation of .29 between the Maudsley Personality Inventory Extraversion Scale and the first reminiscence score, but an insignificant correlation of .10 with a second reminiscence score. In a personal communication, Eysenck indicated that one exact replication and one replication using a different task have given similar significant results. On the other hand, a study using 90-sec. performance trials before rest (rather than five-minute trials) failed to show the expected relationships.

Eysenck has also referred to a conditioning study by Franks (1956) as supporting his hypothesis. Franks found that dysthymics conditioned more readily than hysterics, using

[1] The author wishes to express his gratitude to the University of Illinois Graduate Research Board and to the United States Public Health Service (Project No. M-1713) for their financial support of this project.

the eyeblink reflex and simultaneously the GSR. Like Eysenck, Franks attributed his results to the greater cortical inhibition in the extrovert. But unlike Eysenck, he made it clear that he used the concept of cortical inhibition to describe a more or less enduring trait. Franks wrote, "hysterics are in a state of cortical inhibition in which dissociation phenomena predominate . . . dysthymics are in a state of cortical excitation. . ." (p. 149). There was no indication in Franks' paper that he regarded cortical inhibition as a function of the number of excitations of neural structures mediating stimulus-response connections. To the contrary, he appeared to hold that individuals vary in their basal levels of cortical inhibition and that this was sufficient to account for his results. In a subsequent paper reporting a repetition of his doctoral study on normal *S*s, Franks (1957) shifted his theoretical ground and adopted Eysenck's *reactive* inhibition explanation to account for his findings. As Rechtschaffen (1958) has pointed out, however, it is difficult to defend a *reactive* inhibition hypothesis with eyelid conditioning data because the intertrial interval is not favorable for the accumulation of reactive inhibition. Even if the intertrial interval were favorable for such accumulation, Franks' acquisition curves fail to reveal any systematic tendency to diverge as should be expected if reactive inhibition were accumulating. In Franks' studies, differences were present on the first few trials and maintained thereafter.

A distinction between *basal* and *reactive* cortical inhibition has been proposed previously by Klein and Krech (1952). However, they chose to talk in terms of concepts which are operationally the inverse of inhibition. Their distinction is stated in terms of basal and reactive cortical conductivity rather than inhibition. It is interesting to note that Klein and Krech assumed a high positive correlation between individual differences in basal cortical conductivity and reactive cortical conductivity. If such a correlation could be demonstrated, then an easier way to integrate Franks' results with those of Eysenck would be available.

From this review, it is apparent that three operationally distinct concepts have been associated with the concept of cortical inhibition in one way or another, satiation, reactive inhibition, and basal cortical inhibition. In designing the experiment to be reported here, measurement procedures were included which would cover each of the three inhibition concepts so that the appropriateness of treating all three kinds of measures under a general cortical inhibition concept could be examined.

The idea of relating personality differences to attributes of the nervous system is very enticing. If Eysenck's suggestive results could be expanded upon and repeated by others, a wide area of behavioral study would be opened to more systematic exploration with the promise of unheard of integrations between neurophysiology, learning theory, and personality theory. However, at the start, a number of questions dampen this hope. Rechtschaffen (1958) has rightfully raised a question about the statistical significance of the kinesthetic aftereffect study. Farber, Spence, and Bechtoldt (1957) have found individual differences in eyelid conditioning related to anxiety and *not* to extraversion-introversion. Meier (1956) has reported a negative relationship between kinesthetic figural aftereffects and reminiscence in motor learning. This finding raises the question of whether Eysenck is justified in subsuming satiation effects and reactive inhibition effects under the same construct. Finally, in evaluating Eysenck's results, one is faced with the ever recurring problem of whether findings applicable to selected groups are generalizable beyond these groups. These were some of the hopes, doubts, and confusions which prompted this study.

The main hypotheses to be examined are: (*a*) that individual differences in satiation effects and reactive inhibition effects are correlated, thus, justifying the more general concept of reactive cortical inhibition; (*b*) that individual differences in basal cortical inhibition are correlated with individual differences in reactive cortical inhibition, as assumed by Klein and Krech (1952); and (*c*) that individual differences in satiation, reactive inhibition, or basal cortical inhibition are related to individual differences in extraversion-introversion.

METHOD

Thirty-two male and 30 female students were selected at random from a pool of nearly 1000 introductory psychology students. The mean age of the *Ss* was 20.2 with a *SD* of 2.0. Father's education averaged 11.8 grades with a sigma of 3.5. Because the experiment required eight hours per *S*, $1.00 was paid for each hour's time in excess of the five hours' time each *S* was required to serve as a part of his course requirements.

Assessment of Extraversion-Introversion

Each *S* was given Guilford's Temperament Schedules (13 factors) and Cattell's Revised 16 PF (Forms A and B). In his work, Eysenck has used two criteria in addition to the hysteric-dysthymic diagnosis. At times he has used Guilford's Rhathymia Scale (R) and at others the Maudsley Extraversion Scale (E), which is scored from Guilford's tests. For this study, Guilford's Rhathymia, Eysenck's Extraversion, and an extraversion score derived from the 16 PF by summing factors A, H, E, and F were considered critical for defining extraversion-introversion.

Assessment of Cortical Inhibition

A variety of experimental tests were designed or selected which would produce scores interpretable as measures of basal cortical inhibition, satiation, or reactive inhibition. Since one must stimulate a *S* to get a basal type measure, such measures would not be entirely free of reactive effects. On the other hand some reactive measures cannot be assumed to be free of basal differences. Usually it is only possible to say that the basal variance is probably larger than the reactive variance for such and such a variable, while the reactive variance is larger than the basal for another measure. Of the eight experimental tests used, two clearly measured reactive inhibition effects (pursuit rotor reminiscence and response alteration), three fitted the definition of a satiation effect (kinesthetic aftereffect, Archimedes spiral, and the Necker cube difference score), while three approximated basal type measures (GSR conditioning, aniseikonic lenses, and CFF). Some exceptions to these general classifications are noted later.

The tests were given in three sessions. The first two-hour testing session included GSR conditioning, pursuit rotor, aniseikonic lenses, and Necker cube tests. The second two-hour session included kinesthetic figural aftereffect, CFF, Archimedes spiral, and alternation behavior tests. The third one-hour session, which occurred one to two weeks later, involved retesting on GSR, pursuit rotor, Necker cube, Archimedes spiral, and kinesthetic figural aftereffect.

In the interest of space, only a brief description of each test device is given below. The detailed testing procedures and rationales can be found in Becker (1959) or may be obtained from the author. However, information concerning reliability and construct validity of the measures will be discussed in detail so that an adequate interpretation of the findings is possible.

Measures of Basal Cortical Inhibition

GSR Conditioning. Because most differences in conditionability are apparent in the first few trials, the writer favors Franks' first interpretation of conditioning measures as reflecting basal differences in cortical inhibition. The method consisted of presenting twice by tape recorder a list of 104 words each separated by 5 sec. The word "repeat" was followed by shock on 12 of 20 trials during the first cycle and by a buzzer on 12 of 20 trials during the second cycle. On retest, the list was repeated for one cycle with shock as the US. A Hunter Model 300 GSR amplifier was used to record maximum resistance drop to each stimulus.

The most frequently used measure of conditioning is the number of test trials on which a conditioned response occurs. However, it was found that this kind of measure was greatly influenced by the general reactivity of the *S* as indicated by the magnitude of his response to the US. A measure which was found to be independent of this influence was the number of trials on which the CR was one-half or greater than the preceding UR. This measure, in addition, compensates for the typical adaptation of the UR and CR over trials and for artifact related to basal resistance level and shock level. By this definition, conditioning is considered stronger the closer the CR approximates the UR. Two generalization measures were also available. One based on temporal generalization and the other based on generalization in terms of sound similarity. These measures were combined to obtain a more reliable generalization measure. The generalization measure was also corrected for reactivity by dividing the average resistance drop to the generalization stimuli by the average resistance drop to the US.[2]

The correlations between conditioning measures are presented in Table 1. These intercorrelations can be interpreted as lower bounds of the retest reliability coefficient. They are lower-bound coefficients because they include errors of measurement plus any changes in factor composition of the scores over successive blocks of trials. The combined first session score has a reliability of .80 when the Spearman-Brown estimation formula is applied and correlates .71 with the conditioning measure derived from the second session.

While the usual learning curve type of construct validity evidence is not possible with GSR data because of the complicating adaptation effects, it is possible to compare the resistance drop to the CS and to the generalization words with the reaction to a control word "constant" (which appeared 11 times per cycle but never adjacent to the US) and to the other filler words which appeared but once on each cycle. Table 2 presents such an analysis. In each case the reported mean reaction is based on a *ratio* of a particular set of responses to the *Ss*' response to the US. These measures are therefore analogous to the conditioning measure discussed above, based on the number of trials on which the CS is one-half or greater than the preceding US.

From Table 2 it is apparent that the reaction to the conditioned stimulus is about seven times greater than

[2] This ratio type measure was found to correlate .93 with the measure based on the number of trials on which CR was ½ or more than the preceding UR.

TABLE 1
CORRELATION AMONG CONDITIONING MEASURES

Variables	1b	2	Generalization
1a Shock series, first session	.67	.66	.51
1b Buzzer series, first session	—	.64	.58
2 Shock series, retest session		—	.58
1 First session combined		.71	.59

TABLE 2
EVIDENCE FOR THE VALIDITY OF THE CONDITIONING MEASURES

Stimulus	Resistance Drop to Indicated Stimulus Relative to Resistance Drop to US	
	Mean	Sigma
Conditioned Stimulus		
Repeat (test trials, first session)	46.6%	
Repeat (test trials, second session)	40.1%	
Average to repeat	43.3%**	22.65
Generalization Stimuli		
Crafty (first session)	13.2%	
Crafty (second session)	27.9%	
Receipt (first session)	11.2%	
Receipt (second session)	17.8%	
Average to crafty and receipt	17.6%**	14.66
Control Stimuli		
Constant (first session)	6.4%	
Constant (second session)	6.2%	
All other words (first session)	6.0%	
All other words (second session)	5.2%	
Average to control stimuli	5.9%	11.50

** Significance of difference from control stimuli mean is beyond the .001 level.

that to the control word *constant* and to all other words in the list. Similarly, the reaction to the generalization stimuli is about three times greater than that to the control stimuli. Both of these differences are highly significant (CR = 11.6 and 4.8 respectively) and are consistent with interpreting the proposed measures as indices of conditioning.

Aniseikonic lenses. Aniseikonic lenses have had their main use in the study of stereoscopic perception. The induced effect of the lenses is to make a table appear to tip up like a drawing board or to make a wall lean toward the observer. Becker (1954) has demonstrated marked individual differences in the induced effect and has related the perceived magnitude of distortion and the time required to see the distortion to inhibited and uninhibited personality characteristics. The hypothesis offered is that extroverts with higher cortical inhibition should be less bound to their past perceptual habits and therefore more likely to shift in the new situation. A clear differential prediction, however, between satiation and basal effects can not be made for it is possible to argue that either might lead to weaker perceptual habits in this case. As in the conditioning case, the author is inclined to favor a basal difference hypothesis.

The apparatus and procedures duplicate those used by Becker (1954). Two scores were computed for later analysis: (*a*) *time delay*, which is the average time taken to report an effect on trials three to six, and (*b*) *estimated tilt*, which is the average tilt reported on trials three to six. Becker (1954) has previously reported a stability coefficient of .70 over a three week period for the time delay score. Using an analysis of variance approach, the reliability (consistency) of both scores was found to be .95.

Critical fusion frequency (CFF). Flicker fusion measures have often been considered relevant to the evaluation of the efficiency of cortical functioning (Landis, 1951). It is possible to predict that a higher basal cortical inhibition level would act to lower the perception of discrete flashes (lower CFF). However, since the organism is being stimulated over a period of time, satiation effects could also be operating. In an attempt to isolate basal and satiation effects, two kinds of flicker measures were taken, one freer of satiation effects than the other. The first was based on the usual adjustment method, wherein the rate of flicker was varied until fusion or flicker was reported. This score, based on the average of eight alternating up-down trials taken after 10-min. practice and suitable dark adaptation, will be referred to as CFF *up-down*. The second score was based on one-second exposures of the stimulus at fixed speeds while the subject merely reported "flicker" or "no flicker." This score will be referred to as CFF *median*, since it was based on the median of the interval bounded by four successive reports of "flicker" and "no flicker." It was assumed that the first measure would reflect basal inhibition and satiation effects, while the second measure would reflect basal effects primarily. The difference between the two measures was assumed to reflect a satiation effect. The apparatus consisted of an episcotister model after that used by McNemar (1941), except that a Compur shutter was mounted in front of the viewing glass to permit control of exposure time.

The correlation between the CFF up-down score and the CFF median score was .58. This should not be too high if the difference score is to have any value. Since previous work (McNemar, 1951) has indicated within sessions reliabilities above .95 for CFF judgments after practice, no further check of this was made. The construct validity evidence for these measures is at best suggestive. The CFF median procedure produced an average CFF which was 2.3 cps higher than that found by the up-down procedure. This difference is statistically significant (CR = 4.1) using the appropriate test for the difference between correlated means and is consistent with the satiation hypothesis offered earlier.

Measures of Satiation

Archimedes spiral. In a personal communication, Eysenck has suggested that satiation effects should *de-*

TABLE 3

CORRELATIONS, MEANS, AND SIGMAS FOR ARCHIMEDES SPIRAL MEASURES

Variables	Correlations			*M*	σ
	1b	1c	2		
1a Two-min. satiation	.60	.72	.28	2.29	3.58
1b One-min. satiation	—	.48	.38	1.90	2.71
1c Three-min. satiation		—	.66	2.68	4.12
2 Two-min. satiation, retest			—	2.82	3.94
1 Average of 1a, 1b, and 1c			.53		

TABLE 4

INTERCORRELATIONS, MEANS, AND SIGMAS FOR NECKER CUBE DIFFERENCE SCORES

Variables	Correlations			*M*	σ
	1a	1b	2		
1a One-min. trials	—	−.10	.23	.16	3.75
1b Three-min. trials		—	.38	1.69*	5.15
2 Two-min. trial, retest			—	−.26	2.70

* Significant at the .01 level.

crease the movement reversal aftereffect produced by looking at a rotating spiral. To measure the *length* of afterimage the writer modified the Archimedes spiral test from its recent use by Price and Deabler (1955). A two-revolution spiral was cut on a 1½ in. cylinder of plexiglass and attached to a phonograph motor within a light box. Lights in the box were controlled by a timing circuit which made it possible to have the lights continuously on or off, or to have them cycle on for one sec., off for three sec. *S* viewed the spiral with monocular tube vision. After appropriate instructions and a practice trial, *S* was given satiation trials of two-min., one-min., and three-min. duration. The length of afterimage score was based on the number of light flashes for which contraction of the spiral was reported before four cumulative failures to report contraction. An additional two-min. satiation trial was given during the retest period. In view of their moderately high intercorrelations, the first session measures were averaged. Table 3 presents the intercorrelations of the spiral afterimage measures, their means, and sigmas. Using the alpha coefficient, the within sessions reliability of the averaged first session measure was found to be .82. The between sessions reliability is tolerable only for the averaged measure and for the three-min. satiation measure. The fact that the mean length of afterimage shows systematic increase with length of satiation is suggestive evidence for construct validity. The mean difference between the one-min. and three-min. satiation periods is significant at the .05 level (CR = 1.68) using a one-tailed test.

Necker cube. McDougall, Cattell, and Eysenck have all tried to relate Necker cube measures to extraversion with little success. Eysenck believes the extravert's tendency to look away from the fixation point is the basis for the lack of consistent findings. The present writer is willing to grant that rate of oscillation of reversible figures is probably a function of a neural speed variable (Cattell, 1957) which is independent of cortical inhibition. However, it is possible to predict from the cortical inhibition hypothesis that extraverts should show a greater slowing down of reversible perspective over a period of continuous stimulation as a function of satiation. To guard against the extravert's alleged tendency to "leave the field," the stimulus was presented in a light box, using monocular tube vision. A typical Necker cube with one side shaded was presented to *S* for two one-min. trials and two three-min. trials. One two-min. trial was given on retest. A constant speed kymograph operated by a button was used by *S* to report reversals. The correlations among the difference scores and the means and sigmas are presented in Table 4. Since *S* was given two one-min. trials and two three-min. trials it was possible to compute reliabilities separately for each of these conditions. The reliability of the average of the two one-min. trial difference scores was found to be .17, and that for the three-min. trials, .42. On the basis of the mean difference data, it is apparent that only the three-min. trials produced a significant decrement as would be expected if a satiation effect were being measured. In view of this and the reliability findings, only the three-min. trial measure was retained for further analysis.

Kinesthetic figural aftereffect. This device is presumably a direct measure of satiation effects. *S* first makes a series of width judgments using only kinesthetic cues, and then, following stimulation of the fingers, the judgments are repeated. Eysenck (1955) gives details of rationale and apparatus. In the present experiment, eight baseline readings were made, alternating up and down from random starting points. Then four test judgments were made following satiation periods of 30, 60, and 120 sec. Four more test judgments were taken after 10 min. rest. A 60 sec. satiation period was used on retest after establishing baseline once more. Rate of satiation was controlled by having *S* move his hand in rhythm with a metronome operating at 126 beats per min. The satiation measure is that used by Eysenck (percentage decrement) and is found by subtracting the average postsatiation judgment from the average baseline judgment and dividing by the average baseline judgment.

Eysenck (1955) has reported reliabilities based on an analysis of variance estimate of *r* ranging from .78 to .94 for decrement scores. However, these are undoubtedly spurious. When the same baseline scores are used to compute two decrement scores which are then intercorrelated, a variety of index correlation is produced which has spurious properties. It becomes possible to obtain a correlation of .50 between two decrement scores when in fact there is no correlation between the test judgments or baseline judgments upon which the decrement scores are based. With the means and variances in the present study, the intercorrelations of the four decrement scores from the first session would be about .40 even if there were no correlations (no reliability) between the test judgments and basal judgments involved.

TABLE 5
KINESTHETIC FIGURAL AFTEREFFECTS

Variables	Correlations, Basal Partialed Out		
	1b	1c	1d
A. Intertrial Reliability Estimates, First Session			
1a Mean test judgment, 30" trial	.56	.55	.27
1b Mean test judgment, 60" trial	—	.70	.38
1c Mean test judgment, 120" trial		—	.54
1d Mean test judgment, 10' delay			—

Variables	Correlations with Retest Decrement Score	*M*	σ
B. Between Sessions Reliability, Means, and Sigmas			
1a Percentage decrement, 30" trial	.05	8.83	10.8
1b Percentage decrement, 60" trial	.21	8.97	12.9
1c Percentage decrement, 120" trial	.10	11.26	14.7
1d Percentage decrement, 10' delay	−.21	8.12	13.8
1 Mean percentage decrement, first session	.04	9.29	11.3
2 Percentage decrement, 60" retest	—	5.96	11.8

In view of this, an attempt was made to gain an estimate of the within-session reliability by computing correlations between the various test judgments with baseline judgment partialed out. The results of this analysis are presented in Part A of Table 5. The pattern of the intercorrelations in Table 5 suggests that whatever the judgments are measuring is changing somewhat from trial to trial. However, there appears to be enough common variance to justify adding the four estimates to gain a single more reliable measure. The alpha coefficient, based on the correlations among the test judgments with basal level partialed out, gives a within sessions reliability estimate of .80 for the four averaged scores. Part B of Table 5 indicates that the retest reliability of the decrement scores is nil. Neither Eysenck nor Klein and Krech have reported retest reliability data. It would appear that a stable individual difference is not being measured. This finding raises a serious question about the adequacy of these figural aftereffect scores for testing Eysenck's hypothesis.

When the means are examined, it becomes apparent that satiation-like effects did occur and that they are of the same order of magnitude as those reported earlier by Klein and Krech (1952) and by Eysenck (1955).[3] These indications coupled with the finding of adequate within session reliability suggest that the present experimental procedures have replicated those of Eysenck as intended.

Measures of Reactive Inhibition

Pursuit Rotor Reminiscence. If, as Hull has argued (1943), reminiscence effects are as function of the dissipation of reactive inhibition (I_R), it follows from Eysenck's hypothesis that extraverts should show more reminiscence.[4] Using a standard pursuit rotor apparatus, each *S* was given up to 20 practice trials of 30 sec. duration, separated by 10 sec. rest. When a male reached a criterion of 14 sec. on target, or a female a criterion of 12 sec., for two consecutive trials, the practice period was ended. The average number of trials to criterion was 13.2 (σ = 4.6). Three males and five females were eliminated for failure to reach criterion. After a 5-min. rest, two 7-min. continuous performance trials were given separated by 5-min. rest periods. The final rest was followed by a 2-min. trial. During retesting, an additional reminiscence score was obtained using a 7-min. performance trial, 5-min. rest, and 2-min. performance. Reminiscence scores were based on the difference between the minute's performance preceding each 5-min. rest and the minute's performance following each 5-min. rest.[5] The correlations among the reminiscence scores are given in Table 6. Scores based on the average of the two first-session scores and the retest score were retained for later analysis. The consistency coefficient for the averaged first session score was found to be .69. The lower bound of the stability coefficient for this averaged score was .44. Significant reminiscence effects were obtained for all three measures, although there appears to be a systematic reduction in reminiscence effects over test trials.

[3] For 30", 60", and 120" satiation periods and the 10 min. delay, Eysenck's means were 5.7, 8.4, 10.4, and 4.2; while those reported by Klein and Krech were approximately 6.0, 8.5, 12.0, and 9.0. In a study completed after this paper was written, Northrup and Becker (in preparation) have demonstrated that judgment artifacts can produce similar decrement effects when no satiation is given, the *S* merely being asked to make repeated judgments after rest periods corresponding to the satiation periods.

[4] Rechtschaffen's (1958) argument that the slower dissipation of reactive inhibition in extroverts leads to a conflicting prediction is not critical since Kimble and Horenstein (1948) have demonstrated that most I_R effects are dissipated within five min. Thus, if five min. rests are used, reminiscence scores should largely reflect I_R.

[5] Eysenck (1956) based his reminiscence score on the difference in performance on the last 30 sec. before rest and the first 10 sec. after rest. This scoring was based on the assumption that the warm-up effect after rest (Ammons, 1947) was due to the extinction of conditioned inhibition ($_sI_R$) and therefore, taking a longer postrest interval would confound I_R and $_sI_R$. Finding no evidence to support this theory and finding it difficult to accept in the first place, a scoring procedure more consistent with Ammons' reasoning (1947) was adopted. Experience with the highly unstable 10 sec. measures in preliminary studies also influenced the decision toward a potentially more reliable measure.

TABLE 6

INTERCORRELATIONS, MEANS, AND SIGMAS FOR REMINISCENCE SCORES

Variables	Correlations		M	σ
	1b	2		
1a First session, first reminiscence score	.52	.41	14.8	6.7
1b First session, second reminiscence score	—	.37	12.1	6.2
2 Retest session reminiscence score		—	8.0	7.2
1 Average of 1a and 1b		.44		

TABLE 7

CORRELATIONS, MEANS, AND SIGMAS FOR ALTERNATION RESPONSE MEASURES

Variables	Correlations		M	σ	CR
	B	C			
A. Percentage alternation responses Part b (random)	.25	.30	52.08	13.5	1.22
B. Percentage alternation responses Part c (runs)	—	.37	63.11	18.02	5.75**
C. Percentage alternation responses Part d (random)		—	54.90	13.3	2.92*
D. Average of Parts b and d	.38	—	53.5	10.8	2.55*

* Significantly different from chance beyond the .01 level, one-tailed test.

** Significantly different from chance beyond the .001 level, one-tailed test.

Alternation Behavior in a DRT Setting. A classical application of the I_R concept has been in the interpretation of alternation behavior. In an attempt to measure individual differences in alternation behavior, a discrimination reaction time apparatus was built which permitted *E* to turn on one of three lights (right, middle, left) which *S* could then turn off with one of two levers (right, left). The middle light could be turned on so that only the right or left lever would turn it off. The right and left lights could be turned off by the right and left levers, respectively. This made it possible for *E* to get any combination or series of responses with the right and left hands and then to see which hand was preferred when the middle light was turned on. To prevent *S* from forming hypotheses about sequences, pseudo ESP instructions were given along with the idea that a random sequence was being used. The conspicuous electric timer kept *S* focused on quick responses. After giving suitable instructions, a prearranged program of lights was presented. The program consisted of five parts: (*a*) 10 trials of right and left lights for initial reaction time measurement, (*b*) a *random* series of 60 right, middle, and left lights, (*c*) 87 trials during which *runs* of right and left lights varying from one to five precede the middle light, (*d*) Part b repeated, and (*e*) Part a repeated. The response alternation score was the percentage of trials on which *S* responded to the middle light with an alternation response. These scores were computed separately for Parts b, c, and d. It was expected that Part c would give the highest percentage of alternations since more I_R would build up during a run in which the same response was made several times. Table 7 presents a confirmation of this expectation along with other relevant reliability and construct validity data. The findings in Table 7 suggest some construct validity, but a longer testing period is clearly needed to increase reliability to an acceptable level. For the subsequent analysis, the alternation scores from Parts b and d were averaged.

Summary of Status of Experimental Measures

Table 8 provides a resume of the reliability and internal validity evidence for the experimental measures retained for further analysis. It should be noted that a good proportion of the experimental measures showed mean variation with variation in stimulus conditions which were consistent with satiation or reactive inhibition theory (construct validity evidence). However, the reliabilities of many of the measures leave much to be desired. There will be need to refer back to this table in evaluating the results of later analyses.

RESULTS

The 17 experimental test measures, 18 experimental background measures, 3 sample background measures, 15 personality questionnaire variables, and 5 random variables were intercorrelated and submitted to a centroid factor analysis (fixed, unit communalities) and then rotated to simple structure using the varimax criterion (Kaiser, 1958).

In considering the results, the *r*s among the experimental measures will be examined first. These *r*s are presented in Table 9. The variables have been ordered to conform with the ordering of the summary given in Table 8, and the signs have been changed so that a positive *r* is consistent with the hypothesis of a general inhibition factor. Examination of Table 9 indicates that there is very little in common among the various experimental measures. Out of 136 possible *r*s, only 20 are significant in the predicted direction at the .05 level using a one-tailed test ($r = .21$). There are 6 significant *r*s in the direction opposite to prediction. If one considers only the *r*s which do not involve retesting with essentially the same measure, only 12 are significant in the expected direction.

The picture is not quite so discouraging when one considers only that submatrix of variables with fairly good reliability and

TABLE 8

SUMMARY OF STATUS OF EXPERIMENTAL MEASURES

	Within Session Reliability	Between Sessions Reliability	Construct Validity Evidence
A. Measures probably reflecting basal cortical inhibition			
1. GSR, conditioning (1)[g]	.80[a]	.71	good
2. GSR, conditioning (2)	.71[b]	.71	good
3. GSR, generalization (1, 2)	.59[b]	na[f]	good
4. Aniseikonic, time delay (1)	.95[a]	.70[c]	na
5. Aniseikonic, estimated tilt (1)	.95[a]	na	na
6. CFF, up-down series (1)	.99[d]	.70[d]	na
7. CFF, median, one sec. measures (1)	.58[b]	na	na
B. Measures of satiation effects			
1. CFF, difference score (1)	na	na	fair
2. Spiral afterimage (1)	.82[e]	.53	fair
3. Spiral afterimage (2)	.53[b]	.53	fair
4. Necker, difference score, three-min. trials (1)	.42[a]	.33[b]	fair
5. Kinesthetic aftereffect decrement (1)	.80[e]	.04	good
6. Kinesthetic aftereffect decrement (2)	na	.04	fair
C. Measures of reactive inhibition			
1. Pursuit rotor reminiscence (1)	.68[a]	.44	good
2. Pursuit rotor reminiscence (2)	.44[b]	.44	good
3. Response alternation, average Part b and d	.38[b]	na	poor
4. Response alternation, average Part c (1)	.38[b]	na	fair

[a] Spearman-Brown corrected.
[b] Lower-bound estimate.
[c] From Becker (1954).
[d] From McNemar (1951) using similar apparatus and procedure.
[e] Cronbach Alpha coefficient.
[f] na stands for "not available".
[g] (1) refers to first testing and (2) refers to retesting.

validity evidence. By dropping from the matrix the kinesthetic aftereffect measures (28, 31) which had zero retest reliability, the two difference score measures (38, 41), and the response alternation measure based on random series (42), it is found that 20 of the 66 *r*s in the remaining matrix are significant in the predicted direction. None is significant in the direction contrary to prediction. Furthermore only 6 out of 66 possible relationships show negative signs, whereas 42 are positive and greater than .10.

These *r*s suggest that there may be some significant relationship among the various measures of basal inhibition, satiation, and reactive inhibition, but the present measures are too unreliable to definitively establish such relationships. Examination of the *r*s in Table 9 suggests that the evidence is strongest for the existence of correlations among the basal inhibition measures, and that satiation and reactive inhibition measures while tending to correlate with the basal measures, do not tend to correlate with each other.

Factor Analysis of Variables

In order to save space the index of variables, the correlation matrix, the centroid factor matrix, and the varimax factor matrix for the 58-variable analysis have been deposited with ADI.[6] However, part of the varimax factor matrix is reported in Table 10, and the salient variables on most of the eight rotated factors are discussed in the subsequent text. Guilford and Cattell personality measures (except Rhathymia) which contribute to second-order anxiety or extraversion factors were omitted from the factor analysis to avoid inflation of loadings due to correlated errors.

The first problem in the interpretation of the factors is to decide which, if any, can be considered basal inhibition, satiation, or reactive inhibition factors. To aid in this decision, the critical variables were examined factor by factor to see to what degree the direction and size of loadings conformed to hypothesis.

It can be seen from Table 10 that Factor 3 shows consistently high loadings in the predicted direction on all of the basal inhibition measures and has loadings in the predicted direction (over .10) on 3 of the 6 satiation measures and on 3 of the 4 reactive inhibition measures. On the basis of sign alone, Factor 3 loads in the predicted direction 14 of the 17 variables. No other factor approaches Factor 3 in goodness of fit to the hypothesis of a

[6] Order Document No. 6315 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 9
CORRELATIONS AMONG EXPERIMENTAL MEASURES OF INHIBITION[a]

Variables	17	18	21	46	47	39	40	41[b]	32	33	38[b]	28[b]	31[b]	24	25	42[b]	43
Basal Inhibition																	
17. GSR (1)	—	71	59	23	26	24	28	13	04	12	04	−05	−29	31	18	−08	12
18. GSR (2)		—	58	13	20	20	12	−04	17	11	20	−06	−34	21	00	02	21
21. GSR, generalization			—	16	14	37	38	12	12	20	01	−11	−24	16	06	11	32
46. Aniseikonic, time				—	31	15	−09	−24	14	22	06	−02	−13	13	13	01	14
47. Aniseikonic, tilt					—	04	01	−02	01	−02	−16	−07	−03	03	−03	−06	12
39. CFF, up-down						—	58	−20	04	16	−11	15	−07	01	08	11	05
40. CFF, median							—	68	19	14	−14	−03	−05	03	08	−19	04
Satiation																	
41. CFF, difference								—	19	02	−08	−18	00	03	03	−33	00
32. Spiral (1)									—	53	06	04	−11	−09	−13	−05	08
33. Spiral (2)										—	13	−12	−06	−12	04	−15	06
38. Necker, difference											—	−15	−05	08	−02	−19	−03
28. Kinesthetic AE (1)												—	04	−07	−06	−01	−02
31. Kinesthetic AE (2)													—	−29	−11	11	04
Reactive Inhibition																	
24. Rotor reminiscence (1)														—	44	−13	01
25. Rotor reminiscence (2)															—	−10	−02
42. Response alternation (random)																—	38
43. Response alternation (runs)																	—

[a] Signs have been changed so that a positive r is consistent with a general inhibition factor. An r of .21 is significant at the .05 level, one-tailed test.

[b] Variables of low retest reliability and/or construct validity.

general inhibition factor. However, Factor 3 is probably best considered a basal inhibition factor until more reliable measurement can indicate the limits of its generality.

Two other factors merit consideration. Factor 1 shows a consistent pattern of signs for all four reactive inhibition measures and also loads two satiation and two basal inhibition measures. As will be seen shortly, Factor 1 turns out to be primarily a sex difference factor and thus in terms of present theory, cannot be considered a function of a variety of inhibition. Factor 8, on the other hand, shows consistent loadings on the more reliable satiation measures. This too must be ruled out as a satiation factor on the basis of later evidence.

In summary, the evaluation of the experimental measures indicates: (*a*) the presence of a basal inhibition factor that may also include some aspects of satiation and reactive inhibition measures; (*b*) a failure to find any factor corresponding to Eysenck's concept of reactive cortical inhibition which includes satiation and reactive inhibition measures; and (*c*) a lack of convincing evidence for the existence of unitary traits fitting the definitions of satiation or reactive inhibition. Before leaving Table 10 to discuss each factor, the reader should note the distribution of factor loadings for the five random variables. Nine out of 40 factor loadings equal or exceed .25, the .05 significance level for r ($N = 62$). If factor loadings were distributed as r, only 2 out of 40 loadings would exceed .24. The high loadings of the random variables should serve to caution the reader against over-interpretation of small factor loadings, and should provide an empirical base for interpreting the significance of the factor loadings.

In the paragraphs to follow, each factor will be discussed briefly. Where necessary, salient variables will be presented in tabular form in such a way that the *description given corresponds to a positive factor loading*. The salient variables on factor 1 are as follows:

48. Female sex	74
11. Guilford M, Feminine interests	74
3. Cattell I, Premsia (feminine)	71
49. Young age	53
4. Cattell M, Alaxia (bohemian)	46
5. Cattell N, Naivete	46
46. Aniseikonic, short time delay	40
20. GSR, smaller UR to shock (2)	35
25. Rotor, more reminiscence (2)	30

None of the loadings on this *sex difference* factor are too surprising. The smaller reaction to shock on the GSR is most likely a function of

TABLE 10

A. FACTOR LOADINGS OF EXPERIMENTAL VARIABLES

Variables	Predicted Pattern A or B		Factor Loadings								
			1	2	3	4	5	6	7	8	H^2
Basal cortical inhibition measures											
17. GSR (1)	−	+	0	0	67	0	+	−22	0	−23	58
18. GSR (2)	−	+	0	+	66	+	+	−	−20	0	56
21. GSR, generalization	−	+	0	+	72	−	0	−	−	0	58
46. Aniseikonic, time	−	+	−40	+	29	0	0	0	−21	+	33
47. Aniseikonic, tilt	+	−	+	0	−42	0	−	−	0	−	25
39. CFF, up-down	+	−	24	0	−48	+	+	0	0	0	34
40. CFF, median	+	−	0	0	−49	+	0	0	−50	+	53
Satiation measures											
41. CFF, difference	−	+	27	0	+	0	+	0	66	−20	59
32. Spiral (1)	−	+	−	36	0	0	22	27	0	−24	36
33. Spiral (2)	−	+	−21	39	+	0	0	+	0	−34	36
38. Necker, difference	+	−	0	−	+	−67	0	+	+	+	55
28. Kinesthetic AE (1)	+	−	+	0	0	+	31	−35	0	0	28
31. Kinesthetic AE (2)	+	−	0	0	28	0	25	−	−	−45	39
Reactive inhibition measures											
24. Rotor reminiscence (1)	+	−	23	+	−	0	0	57	0	+	45
25. Rotor reminiscence (2)	+	−	30	0	0	+	24	51	−30	0	52
42. Response alternation (random)	+	−	+	36	−	0	+	−23	38	−	41
43. Response alternation (runs)	+	−	+	26	−48	0	0	−26	0	0	39
Number of loadings over .25:											
Agreeing with Pattern A			2	2	1	0	1	2	3	1	
Agreeing with Pattern B			1	2	8	1	0	3	1	1	

B. FACTOR LOADINGS OF EXTRAVERSION, ANXIETY, AND RANDOM VARIABLES

8. Rhathymia	−08	01	02	08	81	07	−06	03	68
12. Extraversion (Eysenck)	06	−06	11	−12	82	06	07	05	71
13. Neuroticism (Eysenck)	10	66	07	19	−28	07	03	−11	58
14. Anxiety (Cattell)	16	72	02	−16	−14	−15	02	−03	62
15. Extraversion (Cattell), based on the sum of factors A, E, F, & H	−02	−08	13	14	68	14	13	14	56
48. Sex, higher score male	−74	06	15	18	−08	−35	06	−08	74
54. Random Variable 1	13	−06	13	−09	−12	46	04	12	29
55. Random Variable 2	−01	39	−16	−23	05	11	−25	10	32
56. Random Variable 3	−25	−15	18	48	−02	22	06	−21	44
57. Random Variable 4	−09	−03	10	10	−18	03	−42	02	24
58. Random Variable 5	25	29	−14	00	31	−16	21	03	34

Note.—A zero in the table indicates a loading less than ±10; a + or − sign indicates a loading of 10 or greater but less than 20.

E's tendency to set easier shock levels for women than for men. The rotor reminiscence loading is probably a function of the poorer motor skill of women and the resulting greater build-up of interference effects or the greater effort required when skill is low. The shorter time delay with aniseikonic lenses is consistent with an earlier finding by the author and implies a need to control for such a sex difference in univariate studies.

Factor 2 has the following salient variables:

14. Cattell second-order, Anxiety	72
13. Eysenck N, Neuroticism	66
9. Guilford Co, Uncooperative	66
10. Guilford Ag, Disagreeable	64
5. Cattell N, Shrewdness	40
33. Spiral, longer afterimage (2)	39
55. Random Variable 2	39
32. Spiral, longer afterimage (1)	36
42. Response alternation, random series (1)	36

This factor is defined primarily by the Cattell and Eysenck Anxiety or Neuroticism scales. The loadings of experimental variables are largely inconsequential. It is suspected that the longer spiral afterimage scores represent mainly a response bias effect, i.e., the tendency for anxious subjects to agree, and therefore to

continue to say that they see an effect even though they are not sure. Although the conditioning measure used is not the *most* appropriate for testing Spence's theory relating anxiety and conditioning (Farber, Spence, & Bechtoldt, 1957; Spence, 1958), it is interesting to note the failure of conditioning measures to load on this factor. The only GSR measure to approach a "significant" loading was UR to shock (.24).[7]

The salient variables loading on Factor 3 are as follows:

21. GSR, greater generalization	72
17. GSR, greater conditioning (1)	67
18. GSR, greater conditioning (2)	66
16. GSR, lower basal (1)	58
40. CFF, higher threshold, median of 1″ measures	49
39. CFF, higher threshold, up-down series	48
43. Response alternation, fewer alternations (run series)	48
47. Aniseikonic, less estimated tilt	42
37. Necker cube, fewer reversals, last 28 sec. (3 min. trials)	31
50. Father's education higher	30
46. Aniseikonic, longer time delay	29

Factor 3, as noted earlier, comes closest to any of the factors to representing a general cortical inhibition factor. From the predominance of conditioning, CFF, and aniseikonic lense loadings it is apparent that this factor reflects primarily a basal inhibition effect, rather than a reactive inhibition or satiation effect. Of greatest importance is the failure to find *any* personality measures associated with this factor.

Factor 4 does not require much elaboration. It loads in the 70s on fast alternating perspective as measured by the 1-min. Necker cube trials and by the first 28 sec. of the 3-min. Necker cube trials. In addition, it loads greater UR to shock on the GSR (.51 and .43). Finally, *Ss* showing the greater neural reactivity suggested by the above loadings are found to get more hours of sleep per night (.44).

[7] A more appropriate test of Spence's theory, in contrast to Eysenck's, would use the magnitude of conditioned response corrected for basal artifact. Such tests also failed to reveal any significant relationships to anxiety measures. However, in a GSR conditioning study carried out subsequent to the writing of this paper, Becker and Matteson (in press) used extreme groups on anxiety and extraversion dimensions and found confirmation for Spence's hypothesis, but failed again to find confirmation for Eysenck's hypothesis.

Factor 5 requires closer scrutiny. The salient variables loading on Factor 5 are as follows:

12. Eysenck E, Extraversion	82
8. Guilford R, Rhathymia	81
15. Cattell Second-order Extraversion	68
7. Guilford T, Thinking Extraversion	61
5. Cattell G, Low Super Ego Strength	39
26. Kinesthetic figural aftereffect, higher baseline (1)	34
20. GSR, greater average UR (2)	34
28. Kinesthetic figural aftereffect, greater percentage decrement (1)	31
58. Random variable 5	31
29. Kinesthetic figural aftereffect, higher baseline (2)	25
31. Kinesthetic figural aftereffect, greater percentage decrement (2)	25
25. Rotor reminiscence (2)	24

Factor 5 is clearly an extraversion factor, loading all of the criterion personality measures. The only critical inhibition variable to load over .30 on this factor is the kinesthetic aftereffect decrement score (1) which was found to have zero retest reliability and to correlate opposite to hypothesis with the other inhibition measures (Table 9). There are several other reasons why this decrement score loading cannot be considered supportive evidence for Eysenck's hypothesis. First, the correlations of the decrement score with extraversion measures are on the order of .12, quite insignificant. Second, one of the random variables loads as highly. Third, the baseline kinesthetic aftereffect measure loads just as high as the decrement score, suggesting the possibility that any found association between extraversion and the decrement score is due solely to initial baseline difference. The present correlations support this latter hypothesis, though a further study is required to clarify it. The reason for reporting the small retest loading of the rotor reminiscence measure will become apparent later.

Factor 6 appears to be a motor skill factor associated with sex differences. It loads the measures of pursuit rotor skill in the 50's and both reminiscence measures negatively, indicating that less reminiscence occurs with a higher level of skill. Future attempts to measure reactive inhibition as an individual difference are going to have to find some way to eliminate this confounding of reactive inhibition with motor skill.

Factor 7 is interesting because of its co-loading of personality and objective test measures. Cattell's Q_1, Radicalism, loads .55 and Guilford's T, Introspective, loads .31. The main loading of objective tests are on higher CFF threshold (one second exposures) and on reaction time measures. On the objective test side this factor is quite similar to Cattell's U.I. (T) 22 which he calls Corticalertia or Nervous, Alert Reactivity.

Finally Factor 8 appears to be the remaining variance of the general tendency to overestimate the size of the standard test object under both baseline and postsatiation test conditions in the kinesthetic aftereffect experiment. The only personality test measure found to load on this factor is Cattell's B (IQ) which loaded .35. It is very difficult to say just what B is measuring in the particular context of personality test items in which it is given. The present writer finds zero retest reliability of parallel form measures of B with college students.

Correlational Analysis

Although factor analysis can frequently reveal relationships not apparent in a correlation matrix, factor analysis is handicapped by the difficulty one has in stating the significance of a factor loading. One approach to this problem was the inclusion of random variables in the factor analysis. Another approach is the direct evaluation of the probable significance of the correlations between the 17 inhibition and three extraversion measures.

Using a one-tailed test at the .05 level, there are only two significant *r*'s between the 17 inhibition measures and the three extraversion measures. Since both of these significant *r*s involve the same inhibition measure (rotor reminiscence, retest) and are with Eysenck's E and R, even these two significant *r*s are not independent. On a statistical basis this is a finding that would occur with a high probability on a chance basis. Thus, on statistical grounds, one is forced to the same conclusion apparent from the factor analysis, namely, that *there is no evidence for a relationship between any kind of inhibition measure* (basal inhibition, satiation, reactive inhibition) *and extraversion.*

DISCUSSION

Methodological Problems

The evidence of the present study indicates that it is particularly difficult to get reliable measures of individual differences in *satiation* and *reactive inhibition* effects. If the within sessions reliabilities for these measures were high and the between sessions reliabilities were low it might be possible to argue that one is dealing with very unstable variables which show considerable within individual fluctuation. However, since the present within-session reliabilities were also low (.50 or less), it is not yet possible to decide whether the problem is one of learning how to get more reliable measures or whether there are no consistent individual differences in satiation and reactive inhibition effects.

The kinesthetic figural aftereffect device has methodological difficulties connected with it which neither Klein and Krech (1952) nor Eysenck (1955) have attempted to take into account. Unlike the others, the present *E* took retest measures on his *S*s. It was found that mean baseline dropped from 79 to 76 (objective equality equals 73). Using a test for the significance of difference between correlated means, this difference is significant beyond the .01 level. This finding suggests a need to find out just how much decrement would occur from simply taking repeated judgments of the comparison stimulus with no satiation involved at all. It remains a distinct possibility that the relationship of satiation to extraversion found by Eysenck with this device is solely a function of greater initial overestimation by extroverts and subsequent approaching of more objective judgment with practice.[8] Another methodological problem with the kinesthetic figural aftereffect device is the method used in scoring a satiation effect. The percentage decrement score is a very complex function involving both the complexities inherent in a difference score and those inherent in a ratio score. A more directly meaningful approach to establishing relationships to such satiation measures would be to use the partial correlation method discussed earlier. Other statistical problems in

[8] In a subsequent study, Northrup and Becker (in preparation) failed to find a relationship between extraversion and baseline judgments or satiation decrements. Also see Footnote 3.

Eysenck's original report using this device (1955) have been adequately covered by Rechtschaffen (1958) and need not be repeated here except to note that proper statistical analysis would not have led to the conclusion that the results reported by Eysenck were statistically significant. Finally, it should be noted that Crutchfield, Woodworth, and Albrecht (1958) found that the amount of satiation effect produced in the kinesthetic figural aftereffect apparatus correlated .52 with the tendency for satiation effects to dissipate over time. This is contrary to Eysenck's hypotheses that extraverts should build up satiation effects more quickly and dissipate them more slowly.

Reactive Inhibition and Extraversion

It will be remembered that the only "significant" correlations between inhibition and extraversion measures were between the retest reminiscence score and both E and R. The *r*s were .21 and .22 respectively (and .19 with Cattell's Extraversion). Also, this retest reminiscence score loaded .24 on the Extraversion Factor. In view of Eysenck's previous replications, perhaps one should not be quite so hasty in attributing the finding to chance. At one point the writer suspected that the relationship might be attributable to sex differences or to motor skill differences since both of these variables correlated highly with the reminiscence score. Partialing out each of these variables failed to have any effect on the found relationships to the extraversion measures. It is also significant that the found relationship was to the second reminiscence measure, where more *S*s would be approaching their learning asymptotes (after an additional 16 min. of practice) and thus the reminiscence score would more likely be reflecting reactive inhibition rather than other factors associated with the learning process. In reviewing this result, consideration should also be given to Rechtschaffen's (1958) finding of an *r* of .18 between R and an inhibition measure based on the difference in performance on a motor task under massed versus distributed conditions. The above mentioned *r* would be significant at the .05 level, using a one-tailed test, if it were the only relationship being considered. However, Rechtschaffen was in the position of comparing a number of relationships and on statistical grounds the findings could be attributed to chance.

The accumulation of a number of such correlations of borderline significance suggests that Eysenck's results may be replicable. Opposing the acceptance of this conclusion are several other facts. First, the response alternation measures (based on runs) which behaved in accordance with reactive inhibition theory failed to relate to reminiscence or extraversion. Rechtschaffen's (1958) reminiscence measure failed to relate to R. In studies preliminary to this one, the writer failed to find significant differences between extreme extroverts and extreme introverts on pursuit rotor measures under massed and distributed conditions. Also, Miller (1957) failed to find any relationship between R and reminiscence or the difference between massed and distributed conditions on a serial-rote learning task. In the face of these facts, it must be concluded that if a relationship between reactive inhibition and extraversion exists, it is probably of such a low magnitude to be hardly worthy of further study.

Other Related Studies

In weighing the evidence for or against the hypotheses under consideration, several other results should also be considered. In the work preliminary to this study which included rotor reminiscence, visual satiation, aniseikonic lenses, and the Necker cube, no significant relationships to R were found. Farber, Spence, and Bechtoldt (1957) failed to find any significant relationships between eyelid conditioning and R. Rechtschaffen (1958) failed to find any relationship between a Kohler-type visual satiation measure and R. Using the kinesthetic figural aftereffect satiation measure, Crutchfield, Woodworth, and Albrecht (1958) report that among officers greater satiation effects are associated with staff ratings of passivity, suggestibility, lack of dominance, and lack of leadership skills. While the magnitude of these relationships are no larger than those found by Eysenck (.20 to .30), it is of note that they are in *exactly the opposite direction* from those which Eysenck's theory would predict. Finally, mention should be made again of the study by Meier

(1956) in which he found a significant negative correlation between the amount of reminiscence in inverted alphabet printing and the amount of kinesthetic aftereffect. (The *r*s in the present study between kinesthetic aftereffect and rotor reminiscence measures were all negative, ranging from −.06 to −.29).

SUMMARY AND CONCLUSIONS

This study set out to examine Eysenck's hypothesis that cortical inhibition is one of the bases for individual differences in extraversion-introversion. From a review of the literature it became apparent that three distinct concepts were involved in this hypothesis in one way or another. First, Eysenck subsumed both satiation effects and reactive inhibition effects under the cortical inhibition concept. Second, it became apparent that Franks' use of cortical inhibition was quite distinct from Eysenck's. Franks used the concept of cortical inhibition to refer to a basal trait-like difference among individuals which was independent of the number of stimulations of a given neural structure. Thus, in effect, it was found that three different hypothetical variables, all going under the name of cortical inhibition, were assumed to be related to individual differences in extraversion-introversion. The purpose of this study was to find out which, if any, of these three constructs (satiation, reactive inhibition, basal cortical inhibition) could be subsumed under a more general construct, and to find out which, if any, of these constructs were related to individual differences in extraversion-introversion.

Sixty-two college students of both sexes randomly selected from an elementary psychology pool of 1000 students were used as *S*s. Consistent with Eysenck's earlier work, criterion measures of extraversion-introversion were derived from the Guilford and Cattell personality inventories. Each *S* participated in five hours of laboratory testing in addition to the personality testing. A variety of experimental tests were designed or selected which would produce scores interpretable as measures of basal cortical inhibition, satiation, or reactive inhibition. While it was not always possible experimentally to separate basal and reactive effects, two of the eight experimental tests used clearly measured reactive inhibition effects (pursuit rotor reminiscence and response alternation), three fitted the definition of a satiation effect (kinesthetic aftereffect, Archimedes spiral, and Necker cube difference score), while three approximated basal type measures (GSR conditioning, aniseikonic lenses, and CFF). Some exceptions to these general classifications are noted in the text. Including retest measures taken for many of the variables a week or two later, 17 experimental measures were considered critical for the evaluation of the hypotheses.

The findings indicated that the concept of basal cortical inhibition as used by Franks appeared as a unitary factor, but was not found to be related to extraversion-introversion. No empirical evidence was found to support Eysenck's assumption that satiation and reactive inhibition form a unitary trait. Satiation and reactive inhibition measures were found to have some common variance with the basal inhibition measures, but they did not covary with each other. There was no evidence to support Eysenck's hypotheses that satiation measures covary with extraversion measures. There is very minimal evidence that some reactive inhibition measures covary with extraversion. However, there is equally compelling evidence to the contrary. If a relationship between reactive inhibition and extraversion exists, it is probably of such a small magnitude as to be practically and theoretically trivial.

REFERENCES

AMMONS, R. B. Acquisition of motor skills: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263–281.

BECKER, W. C. Perceptual rigidity as measured by aniseikonic lenses. *J. abnorm. soc. Psychol.*, 1954, **49**, 419–422.

BECKER, W. C. Cortical inhibition and extraversion-introversion Unpublished final rep., *USPHS Project No.* M-1713, Univer. Illinois, 1959.

BECKER, W. C., & MATTESON, H. H. GSR conditioning anxiety, and extraversion. *J. abnorm. soc. Psychol.*, in press.

CATTELL, R. B. *Personality and motivation structure and measurement.* New York: World Book, 1957.

CRUTCHFIELD, R. S., WOODWORTH, D. G., & ALBRECHT, RUTH E. Perceptual performance and the effective person. *USAF tech. Note*, 1958, No. 58–60.

EYSENCK, H. J. Cortical inhibition, figural aftereffect, and theory of personality. *J. abnorm. soc. Psychol.*, 1955, **51**, 94–106.

EYSENCK, H. J. Reminiscence, drive, and personality

theory. *J. abnorm. soc. Psychol.*, 1956, **53**, 328–333.

FARBER, I. E., SPENCE, K. W., & BECHTOLDT, H. P. Emotionality, introversion-extraversion, and conditioning. Paper read at Midwestern Psychological Association, Chicago, May, 1957.

FRANKS, C. M. Conditioning and personality: A study of normal and neurotic subjects. *J. abnorm. soc. Psychol.*, 1956, **2**, 143–149.

FRANKS, C. M. Personality factors and the rate of conditioning. *Brit. J. Psychol.*, 1957, **48**, 119–126.

HULL, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.

KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, **23**, 187–200.

KIMBLE, G. A., & HORENSTEIN, B. R. Reminiscence in motor learning as a function of length of interpolated rest. *J. exp. Psychol.*, 1948, **38**, 239–244.

KLEIN, G. S., & KEECH, D. Cortical conductivity in the brain-injured. *J. Pers.*, 1952, **21**, 118–148.

KÖHLER, W., & WALLACH, H. Figural after-effects: An investigation of visual processes. *Proc. Amer. Phil. Soc.*, 1944, **88**, 269–357.

LANDIS, C. Something about flicker-fusion. *Scient. Mon.*, 1951, **73**, 308–314.

MCNEMAR, OLGA W. The ordering of individuals in critical flicker frequency under different measurement conditions. *J. Psychol.*, 1951, **32**, 3–24.

MEIER, M. J. Interrelationships among personality variables, kinesthetic figural aftereffects, and reminiscence in motor learning. Unpublished doctoral dissertation. Univer. Wisconsin, 1956.

MILLER, PAULA E. Reactive inhibition in extraversion and introversion. Unpublished bachelor's thesis, Univer. Illinois, 1957.

PRICE, A. C., & DEABLER, H. L. Diagnosis of organicity by means of spiral aftereffects. *J. consult. Psychol.*, 1955, **19**, 299–302.

RECHTSCHAFFEN, A. Neural satiation, reactive inhibition, and introversion-extraversion. *J. abnorm. soc. Psychol.*, 1958, **57**, 283–291.

SPENCE, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

(Received June 19, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 67-72

CHANGES IN SELF-IDEAL DISCREPANCY IN REPRESSORS AND SENSITIZERS[1]

JOHN ALTROCCHI, OSCAR A. PARSONS,[2] AND HILDA DICKOFF
Duke University

RESEARCH by Rogers and his colleagues has suggested that the concepts of self and ideal-self become less discrepant as a function of brief psychotherapy and that such changes reflect improvement in adjustment (Butler & Haigh, 1954). Other investigators, such as Block and Thomas (1955) and Hatfield (1958), have found evidence of a curvilinear relationship between psychiatric adjustment or personality integration and the discrepancy between self and ideal-self. The latter investigators also demonstrated that persons with very low self-ideal discrepancies tend to use repressive and denying defenses. The purpose of this study is to attempt further specification of the variables related to self-ideal discrepancies by investigating whether people with different defenses or modes of adaptation differ in self-ideal discrepancy and change in different ways subsequent to psychotherapeutic interpersonal interaction.

The mode-of-adaptation dimension which is suggested is repression-sensitization, the terms having been applied by Gordon (1957) to a dimension suggested by Eriksen and his colleagues (Lazarus, Eriksen, & Fonda, 1951) from research in perceptual defense. Repressors are defined as those who tend to use avoidance, denial, and repression of potential threat and conflict as a primary mode of adaptation. Sensitizers are defined as those who are alerted to and perhaps overinterpret potential threat and conflict and who use intellectual and obsessive defenses as a primary mode of adaptation. It has been shown that people are consistent in their use of either repression or sensitization across task situations (Lazarus et al., 1951; Eriksen, 1952) and that repressors, compared to sensitizers, recall more completed than incompleted tasks under threat (Eriksen, 1954), avoid the expression of unacceptable needs on projective tests and have higher perceptual thresholds for material related to such needs (Eriksen, 1951), recall fewer disturbing words (Carlson, 1954), forget in response to implied failure in a learning task (Truax, 1957), and predict that partners' responses will be more similar to their own responses on a personal inventory (Gordon, 1957).

Since repressors tend to remain unaware of threat, conflict, and negative attitudes and to focus on the positive qualities of themselves and others (Leary, 1957), they can be expected to manifest a small discrepancy between self and ideal-self. Since sensitizers are attuned to and tend to ruminate about threat, conflict, and the negative qualities of themselves and others (Gordon, 1959), they can be expected to manifest a high self-ideal discrepancy. Therefore:

Hypothesis 1—Repressors manifest smaller self-ideal discrepancies than sensitizers.

Brief psychotherapy or a similar kind of experience aimed toward personality growth and increased insight into one's behavior with others should cause the repressors to begin to recognize negative aspects of themselves; their self-ideal discrepancies should therefore increase. Such an experience should cause the sensitizers to begin to blame themselves less and to focus less on their negative and more on the positive qualities, and their self-ideal discrepancies should decrease. Therefore:

Hypothesis 2—As a result of training focusing on psychotherapeutic interpersonal interaction, the self-ideal discrepancies of repressors increase and the discrepancies of sensitizers decrease.

[1] An earlier version of this paper was presented at the meetings of the APA in August, 1958. This study was carried out within the context of and was partially supported by Grant M-1599 from the National Institute of Mental Health. Data analysis was also partially supported by grants from the Duke University Research Council. The authors wish to express their gratitude to Gerald Levin and Tallulah Brown for some of their work in the earlier phases of this study and to Robert Dovenmuehle, Faye Spring, and Betty Sue Johnson, co-investigators in the larger project, for their cooperation.

[2] Now at the University of Oklahoma Medical Center.

Subjects in general, however, disregarding their place on the repression-sensitization dimension, should show the effect found by Rogers and his colleagues (Butler & Haigh, 1954), a decrease in self-ideal discrepancy. Therefore:

Hypothesis 3—As a result of training focusing on psychotherapeutic interpersonal interaction, subjects' self-ideal discrepancies decrease.

METHOD

Eighty-eight senior student nurses at the Duke University School of Nursing in 1956–57 composed the experimental (E) group, and 64 senior student nurses in 1957–58 made up a cross-validation (CV) group. *Ss* were tested at the beginning and end of a 16-week period of training in psychiatric nursing. This intensive training was interpersonally and psychodynamically oriented and included individual counseling aimed at helping the student become more aware of the dynamics underlying her own behavior in her relationships with patients. The assumption was made that such training tends to facilitate emotional growth and enhance self-insight in somewhat the same manner as brief psychotherapy. This assumption is supported by the subjective reports of these student nurses and their instructors, by anecdotal evidence in clinical psychology and related fields, and by some indirect research evidence. For instance, Bernstein, Brophy, McCarthy, and Roepe (1954) found that nurses undergoing a similar experience shifted in their attitudes toward interpersonal relationships.

An attempt was made to develop a more detailed measure of the independent variable, the repression-sensitization dimension. For repressors, the most recent studies used the Denial subscale (*Dn*) rather than the whole *Hy* scale of the MMPI (Gordon, 1959; Hatfield, 1958). Because of the shortness of the *Dn* scale and to be consistent with Gordon (1957), the Lie (*L*) and Defensiveness (*K*) scales were also used to select repressors.[3] The intercorrelations of *L*, *K*, and *Dn* range between .32 and .78 (Little & Fisher, 1958; Hatfield, 1958). For sensitizers, the *Pt* scale has traditionally been used. Since *K* and *Pt* correlate negatively, however (Hatfield, 1958), *Pt* was used without the addition of *K*. Again, to lessen the likelihood of measurement error, the Depression (*D*) and Anxiety (*A*) (Welsh, 1956) scales also were used. Thus, the *T* scores of *D*, *Pt* (without *K*), and *A* were subtracted from the *T* scores on *L*, *K*, and *Dn*.[4] In each group, the 15 *Ss* with the highest positive scores were called repressors and the 15 with the highest negative scores were called sensitizers. For comparison purposes, a middle group was included.

Self-ideal discrepancy was assessed by the use of the Interpersonal Check List (ICL) (Leary, 1957), which consists of 128 phrases to be checked true or false, 16 phrases falling into each of 8 octants, each reflecting a different kind of interpersonal behavior. Discrepancies were calculated by adding across octants the absolute discrepancies between self and ideal-self.

TABLE 1
MEAN SELF-IDEAL DISCREPANCIES AND CHANGES IN THE E GROUP

	Pre-training	Post-training	Change	*t* for Related Means	*p*
Repressors ($N = 15$)	24.7	27.5	−2.8	0.7	*NS*
Midgroup ($N = 15$)	36.1	32.1	4.0	1.7	*NS*
Sensitizers ($N = 15$)	54.0	54.3	−0.3	0.2	*NS*
Total Group ($N = 88$)	40.6	37.3	3.3	2.0	$<.05$

RESULTS

The distributions of repression-sensitization scores in the E and CV groups were investigated. Repressors in the E group had scores ranging from +49 to +95 with a mean of +67. Repressors in the CV group ranged from +40 to +80 with a mean of +54 (difference significant, $p < .01$, *U* test). For sensitizers the scores were as follows: E group, −11 to −78 with a mean of −37; CV group, −18 to −116, with a mean of −51 (difference not significant). On this dimension, then, the two groups were not identical.

Tables 1, 2, and 3 show that Hypothesis 1 was clearly supported.[5] Repressors had much smaller self-ideal discrepancies than sensitizers in the E group and in the CV group. Furthermore, this difference was still apparent after training. In every case, the middle group fell between the repressors and sensitizers.

Hypothesis 2, that repressors and sensitizers change differentially in self-ideal discrepancy after training, was not supported in either

[3] Welsh's Repression (*R*) scale was not used because the construction and correlates of this scale (Welsh, 1956) suggest a variable different from the repression referred to here.

[4] The investigators wish to express their gratitude to G. S. Welsh, W. G. Dahlstrom, and H. G. Gough for their advice concerning the measurement of repression-sensitization.

[5] The lower scores in Table 2 result from shifting from Leary's original weighting system to his later simplified scoring, where equal credit is given to every item. Since no published comparison of the two weighting systems was found, the investigators examined the relationships in a small sample and found correlations of .90 to .95.

group (See Tables 1 and 2). There did not, in fact, seem to be any consistency to the changes of repressors or sensitizers.

Hypothesis 3 received partial support. The self-ideal discrepancy of the E group decreased significantly, but a similar trend did not reach an acceptable level of significance in the CV group (See Tables 1 and 2).

TABLE 2

MEAN SELF-IDEAL DISCREPANCIES AND CHANGES IN THE CV GROUP[a]

	Pre-training	Post-training	Change	t for Related Means	p
Repressors (N = 15)	16.0	13.9	2.1	1.0	*NS*
Midgroup (N = 15)	19.9	18.6	1.3	0.7	*NS*
Sensitizers (N = 15)	22.2	21.7	0.5	0.3	*NS*
Total Group (N = 64)	18.4	17.3	1.1	1.4	*NS*

TABLE 3

TESTS FOR DIFFERENCES IN SELF-IDEAL DISCREPANCIES PRE- AND POSTTRAINING

	E Group				CV Group			
	Pre-training		Post-training		Pre-training		Post-training	
	t	p	t	p	t	p	t	p
Repressors vs. Sensitizers	5.16	<.001	4.12	<.001	2.59	<.05	2.63	<.02
Repressors vs. Midgroup	2.40	<.05	.94	*NS*	1.72	*NS*	2.03	<.10
Midgroup vs. Sensitizers	2.81	<.02	3.17	<.01	.83	*NS*	.84	*NS*

In order to understand better how repressors and sensitizers differ in self-ideal discrepancy and to discover whether any reliable changes occurred during this training, a number of further analyses were undertaken. To delineate the components of the highly significant differences in self-ideal discrepancy, octant scores were analyzed. No significant differences were found between repressors and sensitizers on any octant for ideal-self. Table 4 shows that the difference between repressors and sensitizers in self-ideal discrepancy was primarily accounted for by some rather specific differences in self-concept. Sensitizers consistently described themselves as more rebellious-distrustful and, with somewhat less consistency, as more aggressive-sadistic and more self-effacing-masochistic.

In order to discover whether any reliable changes occurred in specific dimensions of self- or ideal-self-concepts and whether any such changes occurred differentially in repressors and sensitizers, analyses of variance of octant scores were attempted. The results suggested that the octant scores were not stable enough for these complex analyses. Therefore, analyses were made of scores that are less specific but presumably more stable—the Love (Lov) and Dominance (Dom) scores, derived by combining octants (Leary, 1957). (See Tables 5 and 6.) The only result which was significant in both groups was that repressors described themselves as more dominant before and after training. This result merely highlights what was revealed earlier (See Table 4): repressors and sensitizers differed in

TABLE 4

PROBABILITY VALUES FROM t TESTS FOR DIFFERENCES BETWEEN REPRESSORS AND SENSITIZERS IN SELF-CONCEPT BY OCTANTS OF THE INTERPERSONAL CHECK LIST, PRE- AND POSTTRAINING[a]

Octant	Description	E Group		CV Group	
		Pretraining	Posttraining	Pretraining	Posttraining
AP	Managerial–Autocratic				
BC	Competitive–Narcissistic				
DE	Aggressive–Sadistic	<.01	<.05		<.10
FG	Rebellious–Distrustful	<.001	<.001	.01	<.01
HI	Self-Effacing–Masochistic	<.01	<.10		<.01
JK	Docile–Dependent				
LM	Cooperative–Overconventional				
NO	Responsible–Hypernormal				

[a] In all significant comparisons, sensitizers had higher scores.

TABLE 5
MEAN LOV AND DOM SCORES FOR SELF AND IDEAL-SELF FOR REPRESSORS AND SENSITIZERS, PRE- AND POSTTRAINING[a]

	E Group		CV Group	
	Pretraining	Post-training	Pre-training	Post-training
Lov				
Repressors	8.03	1.33	5.67	3.79
	(15.97)[b]	(11.01)	(8.91)	(5.99)
Sensitizers	−7.01	−8.77	1.31	.55
	(16.65)	(14.30)	(6.41)	(6.65)
Dom				
Repressors	9.75	9.37	0.96	3.63
	(18.72)	(18.62)	(8.89)	(9.47)
Sensitizers	−2.49	0.26	−2.38	−1.50
	(19.75)	(19.13)	(9.67)	(9.80)

[a] The lower scores for the CV group again result from the use of the newer scoring system. A rescoring of the E group data and a repeat of the analyses demonstrated that the different results for the E and CV groups cannot be accounted for by the change in scoring systems.
[b] Ideal-Self scores are in parentheses.

TABLE 6
F TESTS FOR ANALYSES OF VARIANCE OF LOV AND DOM SCORES

	Self E	Self CV	Ideal E	Ideal CV
Lov				
Between Groups (repressors-sensitizers)	<.05			
Between Treatments (pre- posttraining)	<.025		<.025	<.10
Interaction (groups × treatments)	<.01			<.025
Dom				
Between Groups (repressors-sensitizers)	<.01	<.025		
Between Treatments (pre- posttraining)		<.05		
Interaction (groups × treatments)				

self-concept but not in ideal-self-concept. One treatment effect was significant in the E group and approached significance in the CV group. The Lov score on ideal-self tended to decrease. Several other treatment effects were significant but without consistency between E and CV groups. Thus, there was equivocal support for the contention that this training led to personality changes. The training, however, for the most part seems to have led to different changes in 1956–57 and 1957–58. More convincing evidence of the effect of the training would be differential effects for repressors and sensitizers, which would be revealed by significant interaction variance. No interaction variance was significant in both groups. The ssignificant interactions in Table 6 indicate that repressors in the E group decreased more on Lov in self-concept and repressors in the CV group decreased more in Lov in the ideal-self-concept.

DISCUSSION

The results of this study help to specify one set of variables related to self-ideal discrepancies by demonstrating that repressors and sensitizers differ in self-ideal discrepancy. In addition, the results add to the growing list of correlates of the repressor-sensitizer dimension. Sensitizers have more negative, i.e., hostile and submissive, self-concepts than repressors and therefore a greater discrepancy between self and ideal-self. This negative evaluation of self may serve the purpose of protecting the sensitizer from criticism by others and increasing the likelihood of support from others. This is congruent with clinical observations of extreme sensitizers—depressed, anxious, and obsessive patients—who emphasize their own weakness and helplessness. The positive evaluation of self by repressors, which may be intended to convince self and others that little or no hostility is present, is consistent with clinical observations of such extreme repressors as hysterics, who emphasize their own affection and interpersonal effectiveness and who vigorously repress and deny negative feelings. These findings seem quite parallel to Eysenck's (1947) findings concerning the differences between neurotic introverts (sensitizers) and neurotic extraverts (repressors). Such observations point to the necessity of considering the personality dimension of repression-sensitization when investigating areas like the self-ideal discrepancy. Gordon (1959), furthermore, has shown that repressors tend to assume similarity between self and partner more frequently than do sensitizers when predicting responses to a personal inventory. The question arises whether such

differential sets may also be found in the perception of others as indicated by description of personality attributes. Another line of investigation which might be fruitful would be to see whether other people see repressors and sensitizers as being as different from each other as repressors' and sensitizers' own self-concepts are.

This study has been less successful in clarifying changes in self-ideal discrepancy. The results mildly supported Rogers' contention that psychotherapeutic experiences tend to result in decreases in self-ideal discrepancy (Butler & Haigh, 1954). Similar results, however, have been found for repeated self- and ideal-self-descriptions within five days (Taylor, 1955). Changes in personality measures (such as the ICL) which result merely from repeating the measure need to be further investigated. Moreover, investigation of the self-ideal discrepancy alone may not reveal separate changes in self-concept and ideal-self-concept.

Without comparison with a control group, no changes in self- or ideal-self-concepts in these *Ss* can be definitively attributed to the training. Moreover, the effects for the E group (1956–57) were somewhat different from the effects for the CV group (1957–58). This may relate to the fact that the instructors and the orientation of the training changed between the two school years. It may also relate to the significant difference between the groups in repression-sensitization. Nevertheless, the significant changes which were found were quite consistent with the goals of the instructors. The instructors attempted to help these students from a somewhat repressive culture look critically and realistically at their own values and behavior as these affected their interactions with patients. In both the E and CV groups the ideal-self tended to decrease in Lov. Since Lov is a score which seems to emphasize the receiving as much as the giving of love, this result suggests that, after training, these student nurses were less extreme in the degree to which they wished their interactions with others to be affection dominated. There were also two differential changes for repressors and sensitizers, both involving repressors' changing on the Lov dimension. These results are consistent with the goals of the instructors and are congruent with the impression during data analysis that repressors showed more consistency in responding to the training than sensitizers, who simply seemed to vary more, both between individuals and over time. This is reminiscent of Smith's (1958) finding that *Ss* with greater self-ideal discrepancy and of Engel's (1959) report that *Ss* with negative self-concepts showed greater instability of self-concept. It may be that the sensitizer group includes two or more subgroups of *Ss* and that breaking down the sample on the basis of more specific dynamics might reveal more consistency in their changes. Nevertheless, if any consistent differential self-concept changes occurred in these repressors and sensitizers, other variables or other instruments are needed to assess these subtle changes and to explain individual differences in them. Thus, the contention that this kind of training leads to personality changes similar to those resulting from brief psychotherapy received only equivocal support and must be studied further.

Finally, investigating self- and ideal-self-concepts separately in this study was more fruitful than investigating self-ideal discrepancies. Therefore, it may be more fruitful to study the components or dimensions of self- and ideal-self-concepts and the changes therein than to study gross self-ideal discrepancies.

SUMMARY

From research in psychotherapy, personality adjustment, and perceptual defense, three hypotheses concerning self-ideal discrepancies in repressors and sensitizers were developed. They were tested on a group of 88 student nurses and cross-validated on another group of 64 student nurses who were tested at the beginning and end of an intensive period of training in psychiatric nursing. Repressors and sensitizers were differentiated by MMPI criteria, and self- and ideal-self-concepts were measured by the Interpersonal Check List.

Hypothesis 1, that repressors would manifest smaller self-ideal discrepancies than sensitizers, was supported in both groups. This difference resulted from the sensitizers' tendency to attribute more negative qualities to themselves. Hypothesis 2, that the self-ideal discrepancies of repressors increase and the

discrepancies of sensitizers decrease, was not confirmed. Hypothesis 3, that subjects' self-ideal discrepancies decrease, was partially supported. Some changes in self or ideal-self, especially on the Love dimension, were found and these changes were consistent with the goals of the training, but the effects of the training from one year to the next were not entirely consistent. Studying the components of the self-concept was found to be more fruitful than studying the gross self-ideal discrepancy.

REFERENCES

BERNSTEIN, L., BROPHY, MARIE L., MCCARTHY, MARY JANE, & ROEPE, RUBY L. Teaching nurse-patient relationships: An experimental study. *Nurs. Res.*, 1954, **3**, 80–84.

BLOCK, J., & THOMAS, H. Is satisfaction with self a measure of adjustment? *J. abnorm. soc. Psychol.*, 1955, **51**, 254–259.

BUTLER, J. M., & HAIGH, G. V. Changes in the relation between self-concepts and ideal concepts consequent upon client-centered counseling. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954. Pp. 55–75.

CARLSON, V. R. Individual differences in the recall of word-association test words. *J. Pers.*, 1954, **23**, 77–87.

ENGEL, MARY. The stability of the self-concept in adolescence. *J. abnorm. soc. Psychol.*, 1959, **58**, 211–215.

ERIKSEN, C. W. Perceptual defense as a function of unacceptable needs. *J. abnorm. soc. Psychol.*, 1951, **46**, 557–564.

ERIKSEN, C. W. Defense against ego-threat in memory and perception. *J. abnorm. soc. Psychol.*, 1952, **47**, 230–235.

ERIKSEN, C. W. Psychological defenses and ego strength in the recall of completed and incompleted tasks. *J. abnorm. soc. Psychol.*, 1954, **49**, 45–50.

EYSENCK, H. J. *Dimensions of personality*. London: Routledge, Paul, 1947.

GORDON, J. R. Interpersonal predictions of repressors and sensitizers. *J. Pers.*, 1957, **25**, 686–698.

GORDON, J. R. The stability of the assumed similarity response set in repressors and sensitizers. *J. Pers.*, 1959, **27**, 362–373.

HATFIELD, J. A study of self-concept configurations in relation to forms of ego functioning. Unpublished doctoral dissertation, Univer. California, 1958.

LAZARUS, R. S., ERIKSEN, C. W., & FONDA, C. P. Personality dynamics and auditory perceptual recognition. *J. Pers.*, 1951, **19**, 471–482.

LEARY, T. *Interpersonal diagnosis of personality*. New York: Ronald, 1957.

LITTLE, K. B., & FISHER, J. Two new experimental scales of the MMPI. *J. consult. Psychol.*, 1958, **22**, 305–306.

SMITH, G. M. Six measures of self-concept discrepancy and instability: Their interrelations, reliability, and relations to other personality measures. *J. consult. Psychol.*, 1958, **22**, 101–112.

TAYLOR, D. M. Changes in the self concept without psychotherapy. *J. consult. Psychol.*, 1955, **19**, 205–209.

TRUAX, C. B. The repression response to implied failure as a function of the hysteria-psychasthenia index. *J. abnorm. soc. Psychol.*, 1957, **55**, 188–193.

WELSH, G. S. Factor dimensions *A* and *R*. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956. Pp. 264–281.

(Received June 30, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1 73-81

EFFECTS OF GROUP SUPPORT ON THE EVALUATION OF AN ANTAGONIST[1]

LLOYD H. STRICKLAND[2]
University of North Carolina

EDWARD E. JONES
Duke University

AND WILLIAM P. SMITH
University of North Carolina

OUR IMPRESSIONS of others are determined not alone by information which these others directly provide; the verbal and gestural responses of a "stimulus person" (*SP*) are typically assessed by a perceiver in relation to the situation in which they occur (Jones & Thibaut, 1958). In many cases, the perceiver himself is a component of this situation. He must therefore consider his own behavior as a condition affecting the other's responses and the meaning he assigns to them. If A, for example, is aware of his own role in the instigation of hostile behavior in B, he is less likely to think of B as hostile or unpleasant, than he would if he were unaware of the provocation he himself provided.

It seems to follow that, in the typical social interaction, evaluations of others on the basis of their responses are contingent on the perceiver's evaluation of his own behavior. If the perceiver behaves in a "good" fashion and this elicits a "bad" response (e.g., a negative evaluation, a frown, a critical comment) from the *SP*, the latter will be judged as personally bad by the perceiver. If the perceiver is unconvinced that his own behavior is good, worthy, or appropriate, he will be less likely to form a negative evaluation of *SP*. This line of reasoning suggests an experimental situation in which one person (the *SP*) responds with hostility to the behavior of another (the perceiver) which is or is not highly valued by the perceiver himself.

A potential determinant of this valuation is the extent to which the perceiver's initiating response has normative support in some relevant group. A hostile reaction to behavior that has such normative support would seem to invite retaliation in the form of a negative impression. Hostile reaction to the same initiating response unsupported by social sanction would result in a less derogatory impression of the hostile respondent.

In simplified form, this is the reasoning behind the main prediction of the experiment described below. Consider an experimental procedure involving three primary steps or stages. In Step I, one group member (*S*) chooses influence messages to transmit to the target person or *SP* (not a group member); in Step II *S* learns that others in the group do (Group Support) or do not (No Support) endorse the particular messages selected; in Step III, *S* hears the *SP* severely criticize his messages and derogate their source. In such a situation, the main experimental hypothesis is that *perceivers in the group support condition will be more negative in their final impressions of the SP than will perceivers in the no support condition.*

Although it is theoretically quite possible that *S*s could respond to such a situation by rejecting their group in the No Support condition or succumbing to various forms of withdrawal and perceptual distortion, the experiment was designed to minimize such alternative reactions and thus to promote differential perceptions of the *SP*. Several post-experimental measures failed to reveal any tendency for *S*s in either group to adopt these different ways of managing the problem created by the *SP*'s hostility. Therefore, these

[1] This study was initiated in the program of the Organization Research Group of the Institute for Research in Social Science at the University of North Carolina. The ORG is supported by the Office of Naval Research Nonr-855(04). Analysis and writing time was made available to Jones on National Science Foundation contract G8857 administered by Duke University.

[2] Now at Dartmouth College.

alternative reactions will not be considered further.

METHOD

Procedure

Ninety male *Ss*, recruited from introductory psychology courses, were observed in groups of three. Each group was told that the experiment was part of a broader research project concerned with opinion change and reactions to propaganda. The experimental rationale emphasized that any propagandist must be aware of the "characteristics of his audience," which could be inferred by the audience's response to an influence attempt. *Ss* were told that one of their group would eventually present a number of persuasive arguments about "big time athletics" to a fourth student (who would be in an adjacent, sound-proof room) while the two remaining *Ss* would be "innocent bystanders." Their job was two-fold: (*a*) evaluate the persuasive arguments to be used; (*b*) observe the reactions of the fourth student and make some inferences about "just what kind of a person you think he is." It was explained to the group that the issue of "big time athletics" on the college campus had been chosen as the influence topic because of the differences in opinion on this subject, and that the *Ss* had been selected to participate in the experiment because they had all had about the same scores on an Attitude-Toward-Big-Time-Athletics questionnaire, which they had filled out in class early in the semester. An attempt actually was made to compose groups of persons whose scores were relatively homogeneous and extreme in the *pro* or *con* direction, although it was not always possible to compose the groups of extremes. The fourth student was presented as one whose questionnaire responses reflected that he was "more or less undecided on the issue".

E stated that he would first interview the fourth student (henceforth referred to as the *SP*) in the next room. This interview would be carried through a microphone to a loudspeaker in the laboratory, so that the three *Ss* might obtain an initial impression of him which would aid in their selection of good arguments supporting their own position. *E* left the room and turned on a previously recorded taped interview.

In this initial interview, an attempt was made to create a rather neutral but mildly likeable *SP*. He was presented as a sophomore, undecided on his major, who enjoyed intramural sports but was concentrating on his studies. He stated that he didn't care one way or the other about "big time" sports—he enjoyed going to ball games but didn't think sports "should push everything out of the picture."

After hearing this brief interview, the *Ss* filled out a 10-item rating scale designed to reveal their "first impressions" of *SP*. They then were asked to bring up in a brief discussion those points (favoring their own position) which might be successful in influencing the interviewee. This was done in order to involve the *Ss* more strongly in the issue and to increase the saliency of group membership by stressing attitudinal similarity and the cooperative nature of the task of finding arguments. Each *S* was then seated in his own private cubicle and assigned a letter (supposedly, *A*, *B* or *C*) for identification. The ostensible purpose of this procedure was to preserve the anonymity of the responsible *S*, whose arguments were to be transmitted to *SP*, vis-à-vis the innocent bystanders, who would be evaluating *S*'s arguments. Actually, all three *Ss* received *B* identification cards.

Each *S* then selected from a list of 13 typical, one-paragraph arguments, the five which he thought most likely to influence *SP*, after having rated each of the 13 on a six-point scale from Poor to Excellent. He made these choices by placing a checkmark in a column headed B, in the left-hand margin of the pages, beside those preferred arguments. *E* then collected these sheets, supposedly to record *S*'s choices and to check in the adjoining A and C columns the choices of *B*'s partners. They were then returned to *S*.

In this way, the major independent variable was manipulated: in the Group Support condition, each *S* (*B*) perceived that *A* and *C* had agreed with his choices in four of five cases. In the No Support condition, *S* found his partners in close agreement with each other but in slight agreement with him.

Following the return of these argument selection sheets to the *Ss*, *B* was formally selected (presumably by lot) as the letter designation of the Communicator, leaving the nonexistent *A* and *C* to function as Innocent Bystanders. All *Ss* were instructed to select, from a pile of hand-written copies of the arguments, the five that *B* chose. *B* was told to rank order them in terms of their probable increasing effectiveness—*A* and *C* were to write brief evaluative comments about them. The rank-ordered arguments were collected from each *S*, and *E* announced that he would bring them to *SP*, let him read them, and then interview him about his reactions to them.

In the next interview, also prerecorded on tape, *SP* appeared initially vague and evasive. As he reviewed each argument (in *S*'s designated order), he became increasingly derogatory, not debating the point made by the argument, but disparaging the attempt to change anyone's mind with such poor material. After completing his review of the arguments, *SP* made a few additional unkind remarks about *S* and then asked *E* if he could leave, preferring not to meet his fellow *Ss*. After granting this permission, *E* returned to the adjoining laboratory and required the *Ss* to complete the following:

1. A brief, free-response personality sketch of *SP*, and a statement of arguments *S* might use were he to attempt to influence *SP* again.
2. A second, private rating of *SP* on the same scale form used in the initial impression rating.
3. A re-rating of all 13 arguments on the original six-point poor–excellent scale.
4. A "cohesiveness" measure, in which *S* designated (*a*) how much he would like to work with his fellow group members again, and (*b*) how much he felt they would prefer him as Communicator again.
5. A "public" rating of *SP*, on a parallel form of the scale used in the first and second ratings, purportedly to serve as the basis for a group discussion about *SP* and final composite group rating of him. It was expected that the predicted differential change in the private ratings would be augmented in making the public ratings.

6. A recognition-recall measure, in which *S* attempted to designate those five arguments initially preferred by himself and his two partners.

After these procedures were completed *E* informed the *S*s about the method and purposes of the experiment and determined the extent to which their suspicions had been aroused.

Rating Scale

As indicated above, each *S* rated the *SP* on three different occasions, before the derogation, after the derogation (private), and once again (public). Since the final two ratings were made consecutively, it was necessary to use two equivalent forms of the scale. In order to insure that the actual positive or negative results obtained were not solely a function of the particular items comprising the scale, approximately half of the *S*s were given Form I to use on the first and second rating and Form II to use on the third (i.e., public) rating. The order of form administration was reversed for the remaining *S*s. Each form of the scale consisted of 5 trait pairs (or 10 items) which measured the perceived characteristics: likeability, adjustment, warmth, intelligence, and conceit. The ratings were made on a six-point scale reflecting the degree of agreement with simple declarative statements about the person being rated. The scale was constructed in such a way that a positive statement like: "He seems to be a very warm and affectionate person" was in all cases matched by its negative counterpart in a different item such as: "He strikes me as a hostile and unsympathetic person." Many of the items on both forms of the scale were taken from a very similar scale described by Jones, Hester, Farina and Davis (1959). Data on reliability are included in the earlier study, but in view of the way in which the scales were used in the present study, separate reliability analyses were run on each scale. The split-half correlations of the change scores on Forms I and II were .57 and .61 respectively (corrected by the Spearman-Brown prophecy formula). Obviously, the scales each measure other variables in addition to overall favorability or "halo," but the reliability coefficients seem adequate to provide a meaningful test of changes in general evaluation of the *SP*.

Design and Distribution of Subjects

Although primary interest lies in the effects of the manipulation of group support, the two additional control variables make the design somewhat more complex. As Table 1 indicates, the design involved three dichotomous variables resulting in eight separate cells of *S*s. Table 1 also indicates the number of *S*s actually observed in each condition. It will be apparent that a large number of *S*s were suspicious before, or became suspicious during the course of, the experiment. *S*s were classified as suspicious if they acknowledged any doubts about the status of the *SP* during the post-experimental group discussion led by *E*. This is a very stringent criterion of suspicion, since some *S*s were probably influenced by statements of suspicion made by others in their group, or their suspicions may have reached a threshold of awareness during the tedium of filling out the final questionnaires. Nevertheless, for most of the analyses to be presented below, all suspicious *S*s by this criterion were excluded. The fact that 29% of the *S*s expressed some degree of suspicion clearly raises some questions about the representativeness of the remaining sample. More will be said about this later, but the main reason for suspicion is easy to locate: the *SP* was forced by the conditions of the experiment to respond with very vague generalities, and to avoid any reference to the content of the arguments he had presumably received. This unavoidable circumstance combined with the surprise and incredibility of the actual derogation of *S*, made it difficult for some *S*s to accept the *SP* as a naive fourth *S*. It should be noted that eight *S*s were discarded because of failure to comply with the instructions (e.g., they sent unpreferred arguments to *SP*), or because they felt their stand on Big-Time Athletics had been misidentified.

RESULTS

Success of Experimental Manipulations

Response to the SP. The script for the tape recorded response to the *S*'s arguments was constructed with the purpose of instigating the *S*s to hostility. Almost without exception, *S*s did in fact change their ratings in a negative direction. Pooling the change scores of all *S*s

TABLE 1
EXPERIMENTAL DESIGN AND DISTRIBUTION OF *S*s

	Group Support					No Support				
	Pro BTA		Anti BTA		TOT	Pro BTA		Anti BTA		TOT
Rating Form Sequence:	I-II	II-I	I-II	II-I		I-II	II-I	I-II	II-I	
Good *S*s	9	6	6	8	29	6	6	8	7	27
Suspicious[a]	7	5	0	2	14[b]	3	0	4	5	12[c]
Total	16	11	6	10	43	9	6	12	12	39

[a] Although suspicious *S*s appear to be unevenly distributed between conditions, a chi square test reveals that this distribution does not, in fact, depart from chance.

[b] None of these *S*s spontaneously expressed suspicion in the free-response impression sketch.

[c] Five of these *S*s did spontaneously express suspicion.

TABLE 2
MEAN BEFORE, AFTER AND CHANGE SCORES, SUMMARY RATING EVALUATIONS

	Group Support					No Support				
	Pro BTA		Anti BTA		Total	Pro BTA		Anti BTA		Total
Form	I	II	I	II		I	II	I	II	
N	9	6	6	8	29	6	6	8	7	27
Before[a]	3.67	3.92	4.22	4.09	*3.95*	4.63	4.00	4.41	4.26	*4.34*
After[a]	2.81	3.03	3.30	3.35	*3.11*	3.78	3.40	3.79	3.96	*3.75*
Change[b]	.86	.89	.92	.74	*.84*	.85	.60	.62	.30	*.59*

[a] The higher the score, the more favorable the overall evaluation.
[b] The higher the score, the greater the change in the negative direction.

and comparing the mean against the null expectation of zero change results in a t value of 16.4, $p < .001$, leaving little doubt that the *SP*'s recorded comments had their intended effect.

Perception of social support. A crucial condition for an adequate test of the major hypothesis is that the degree of support for the *S*'s arguments, conveyed to him through an exchange of bogus ballots, should be actually perceived as high or low in line with the degree of agreement indicated. Suffice it to say that *S*s in the two conditions are strikingly different from each other ($p < .001$) in amount of agreement perceived (on the recognition–recall measure).

It is possible, of course, that agreement perceived has no implications beyond the purely cognitive ability to recall the bogus ballots distributed. To what extent does each *S* actually infer that the group is behind him, and accepts him as spokesman? Does this inference vary with support treatment? On the postexperimental questionnaire the *S*s were asked to predict whether the other two members of the group would like to have them as "communicator" again. By classifying the responses as yes or no, a chi square test was computed indicating a clear association between support condition and perceived acceptance ($\chi^2 = 5.22$, $df = 1$, $p < .025$). Thus, *S*s in the No Support condition felt that the group was unhappy with their performance and would not be likely to support them in the same role again.

The Effects of Variation in Group Support

Summary rating evaluation. The main hypothesis of the present study was that *S*s in the Group Support condition would show a greater increase in negative evaluation of the *SP* than would *S*s in the No Support condition. This prediction was expected to hold regardless of variation in such control variables as attitude toward athletics and rating form used. Before examining the evidence concerning this main hypothesis, it may be pointed out that neither of these control variables was seen to have any effects of significance in analyzing the main rating data. For presentational simplicity, therefore, no further reference will be made to these variables unless they are specifically relevant to a particular point of difference.

The data most clearly relevant to the main hypothesis are presented in Table 2 and analyzed in Table 3. Here it may be seen that the difference between Group Support and No Support conditions is consistently in the predicted direction for Before, After, and Change scores. The difference between After scores of the two treatment groups is highly significant, the difference between Before scores is significant at the 5% level, and the crucial difference between Change scores does not approach significance.

Interpretation of this pattern of findings is complicated by the quite unexpected difference in Before scores. This result requires explanation in and of itself, because *S*s were randomly assigned to the two treatment groups and at the time the Before measure was taken, *S*s had not been exposed to any differential treatment. The apparent explanation is a curious and unusual one: *S*s in the Group Support condition who initially liked the *SP* tended to become suspicious that he was not responding spontaneously and were later discarded as *S*s; *S*s in the no support condition

TABLE 3
SUMMARY OF ANALYSIS OF VARIANCE[a] RESULTS
(Summary rating evaluations: Before, After, and Change scores)

Source	df	Before		After		Change	
		MS	F	MS	F	MS	F
Support (A)	1	.25	4.17[b]	.75	9.38[c]	.12	2.4
Attitude (B)	1	.06		.24		.05	
Form (C)	1	.05		.00		.05	
A × B	1	.07		.00		.02	
A × C	1	.10		.03		.01	
B × C	1	.00		.02		.01	
A × B × C	1	.09		.06		.01	
Error	48	.06		.08		.05	

[a] Because of the unequal cell frequencies, the approximation technique presented by Snedecor (1946) was used.
[b] $p < .05$.
[c] $p < .005$.

who initially disliked the *SP* also tended to become suspicious. The net result of this interaction between support treatment and initial impression, as it affects arousing suspicion, is that usable, nonsuspicious *S*s have significantly different Before scores. The causes of this differential suspicion will be discussed below, but the consequence of the systematic variation in Before scores is the major concern in the present context. It is difficult to interpret the Change scores at face value when the group which is predicted to show the greater decline in rating starts out significantly lower. Both the "floor" effect and the tendency for regression toward the mean would tend to suppress the Change scores in the group support treatment, where the greatest change is predicted. Such a regression effect would be especially likely, given the moderate reliability of the Change scores.

There turns out to be no simple or single way of handling this problem. Two bits of evidence do tend to clarify the picture, however, in the direction of partial support for the major hypothesis of difference between Group Support and No Support conditions. First of all, if the Before scores for the two main treatment conditions are matched by discarding 11 *S*s, a t test for matched groups shows the Group Support and No Support conditions to be significantly different ($p < .05$, one-tailed test). This is, of course, a rather wasteful procedure and one which limits the ease of generalization to a broader population of college students.

Secondly, looking at the regression of After scores on Before scores, the coefficient for the No Support *S*s ($b = 1.09$) is significantly greater than the coefficient for the Group Support *S*s ($b = .42$), the probability value of the difference being less than .01. Examination of the scatter plots for the two conditions indicates that this difference is mainly a function of the initially favorable *S*s, who show much more marked decline in the Group Support than in the No Support condition.

Support for this last notion appears when the Before scores for the two treatments are split at the overall median. The difference between Before score means for the two high subgroups is not significant, and these *S*s, who are initially favorable to the *SP*, show significantly greater negative change ($p < .01$) in the Group Support condition than in the No support condition.

With the low subgroups, *S*s in the Group Support condition had significantly lower before scores than those in the Support condition ($p < .01$) but not significantly greater Change scores. Thus, the problem of interpretation concerns only those *S*s who were initially unfavorable; the results for the initially favorable *S*s are straightforward.

To summarize, then, the hypothesis that evaluative ratings will become more negative in the Group Support than in the No Support condition receives only qualified support. A more accurate statement of the general hypothesis would be: those *S*s who show an initially favorable impression of a person who later derogates them for their action will show a greater decline in evaluation when their actions have received normative support than when they have not.

Free response descriptions. In addition to the summary evaluation provided by averaging the item scores on the rating scale, it is important to consider any support for the main hypothesis derived from analysis of the impression sketches contributed by each *S* immediately after the derogation of his arguments. Although the sketches were rich in variety, most of the *S*s seemed concerned with "explaining" the *SP*'s behavior, either by referring to his rigidity, stupidity, and inflexibility, or to his honesty and sense of independence.

Before discussing the effects of the group

support variable on these free response descriptions, it is necessary to consider once again the related problems of sampling and suspicion. Since the impression sketches were only written once, after the *S*s had heard both parts of the tape recording, it is obviously impossible to obtain a measure of change from a "before score" baseline. But the significant difference in Before score on the ratings of nonsuspicious *S*s has already been noted, thus tending to confound any differences in the same direction which might be observed on the "after" impression sketches. As a way of handling this problem, the sample was reconstituted for the purpose of this particular analysis. From the total pool of *S*s, only those were discarded who showed any signs of suspicion in the impression sketch itself, although some of them later reported varying degrees of suspicion under questioning. The remaining sample contained 43 *S*s in the Group Support condition and 34 *S*s in the No Support condition. The primary reason for extending the sample in this way was to assure greater comparability between the Before scores (rating scale summaries) of the Group Support and No Support *S*s. This purpose was accomplished since the difference in Before scores between the extended subsamples did not approach significance. We are presumably free, then, to interpret any difference between support conditions on the impression sketches as largely the result of experimental manipulations rather than sampling variation.

Each of the impression sketches was placed by two coders in one of five carefully described rating categories arrayed on the dimension of overall favorability. As evidence of the reliability of coding, 58% of the placements were identical, and 98% of the placements fell in identical or adjacent categories. In cases of disagreement, the final score assigned was the average of the two ratings.

The final distribution of ratings was quite bimodal and skewed, with more of the impression sketches being clearly negative in tone. For this reason, the scores were split as near to the median as possible and the resulting two by two table (More Favorable–Less Favorable, Group Support–No Support) was submitted to a chi square test. The obtained value of 11.97 (corrected for continuity, $df = 1$) is significant at the .001 level, indicating a strong association between support condition and the expression of hostility—those in the Group Support condition express more hostility in their impression sketches than those in the No Support condition.

In general, the Group Support *S*s were more likely than the No Support *S*s to refer to the *SP* as prejudiced, ignorant, egotistical, indifferent, or confused. The No Support *S*s, on the other hand, were more likely to refer to the *SP* as likeable, rational, and honest.

Publicity of Ratings and Group Support

As indicated above, a comparison of the change in summary rating evaluations for Group Support and No Support *S*s provides qualified support for the main hypothesis. That is, the Group Support *S*s show greater negative change than do the No Support *S*s. To what extent does this reflect a private change in judgment, and to what extent are we dealing with differences in the inhibition versus facilitation of the *expression* of hostility? It might be argued that at least some of the *S*s held the vague expectation that they would have to justify their ratings to their fellow group members. If this were the case, then these *S*s would feel more safety in expressing their hostility on the rating scale in the Group Support than in the No Support condition.

No one has yet discovered a simple way of measuring the difference between private feelings and public expression of these feelings. Some form of expression is inevitably bound up with the measurement of attitudes or impressions. However, some partial evidence relevant to this point is provided in the design of the present experiment. It will be recalled that the *S*s filled out two forms of the rating scale after the derogation. First they rated the *SP* on one 10-item form under instructions which stressed the anonymity of the ratings and the fact that they would only be seen by *E*. This "private" After scale was always the same form as the Before scale. Then they were asked to fill out a different 10-item scale under the explicit instructions that each member's ratings would serve as the basis for a three-person group discussion, a discussion which would lead eventually to a common rating of the *SP*. If the observed changes in rating reported above were entirely a function of inhibition or facilitation of reporting one's

hostility, there should be no difference between private and public ratings. If social inhibition and facilitation exercise independent effects, we should expect the difference between Group Support and no support conditions to be even greater on the public ratings.

There are several kinds of analysis that would be appropriate in answering this question, but perhaps the most straightforward involves tabulating the number of *S*s who become more negative and the number of *S*s who become less negative going from the private to public rating form in the Support and No Support conditions. By omitting the few *S*s who did not change in either direction, the resulting two by two table shows that 14 out of 25 Group Support *S*s were more negative in their public ratings, whereas only 6 out of 25 No Support *S*s became more negative and 19 more positive. This difference is significant by chi square test at the .05 level. In other words, when faced with the prospect of publicizing and defending their ratings to the other two group members, the Group Support *S*s became somewhat more hostile on the average. When faced with the same prospect, the No Support *S*s became considerably less hostile on the average.

The interpretation of this finding is complicated by the fact that almost all of this difference is attributable to the group of *S*s rating the *SP* privately on Form II and publicly on Form I. Those *S*s going from Form I to Form II show a slight and nonsignificant change in the same direction. Since there are no differences between Form I and Form II means for any other comparison (Before ratings, Change ratings, private After ratings) it is difficult to know what to make of this variation. However, since the overall effect is significant when both forms are pooled (both by chi square and *t* test) it seems reasonable to state the guarded conclusion that there are independent inhibitory effects stemming from the anticipation of group confrontation in the No Support condition and (to a lesser extent) facilitating effects in the Group Support condition. This does not mean that the private ratings were entirely unaffected by such factors, but it does suggest that they largely reflect the *S*'s internalized impression of the *SP*.

Suspicion as an Alternative to Changes in Evaluation

As noted above, there was no reliable tendency for any of the *S*s to reject the group, devaluate the arguments, or misperceive the degree of support in order to maintain their initial impression of the *SP*. It is possible, however, to consider suspicion as one form of alternative reaction to the experimental situation. As already reported, *S*s in the Group Support condition who initially liked the *SP* tended to become suspicious, whereas, those in the No Support condition who initially *disliked* the *SP* tended to become suspicious. What is the psychological significance of becoming suspicious in the present context? Presumably, *S*s in a psychological experiment become suspicious because (*a*) the situation is poorly rationalized by the *E*, there are incongruous features which destroy credibility, etc., or (*b*) *S* desperately wants to believe that the situation is not as described. In most cases of suspicion these two factors probably reinforce each other—especially when something unpleasant or stressful is done to the *S*. In the present experiment two kinds of *S*s became suspicious. Let us examine each of these in turn.

Considering first those *S*s in the Group Support condition who initially liked the *SP*, it may simply be hard for them to believe that a person for whom they have esteem and respect is responding with such antagonism to so little provocation. The incredulousness of these *S*s may be heightened by their awareness that two other *S*s have checked almost the identical five arguments which were sent. We may speculate, then, that these *S*s became suspicious primarily because of the cognitive–affective incongruity of the experimental situation.

As for those *S*s in the No Support condition who dislike the *SP* initially, the unpleasantness of their position should be noted. They dislike the *SP* and the *SP* appears to dislike them. And yet, the other two group members have indicated that they would have sent different arguments had either of them been communicators. In this situation of unrelieved antagonism and lack of support, it may be that the *S* is highly motivated to perceive any cues of artificiality or incongruity.

The fact that different kinds of people may have become suspicious in the two conditions raises the issue of confounding in the rating results. However, including most of the suspicious *S*s (as in the free response data) results in even greater differences in the predicted direction. Furthermore, there is no a priori reason for assuming that suspicion would consistently bias the results to favor confirmation of the hypothesis.

DISCUSSION

The present research grew out of a general concern with the conditions of interaction which affect the perception and evaluation of a stimulus person. It is to be noted that the behavior of the person being judged was constant and identical from *S* to *S*. The meaning of this behavior, however, was shown to vary with the interaction context from which it stemmed. In the experiment, a stimulus person responded in an abusive, antagonistic fashion to a series of arguments chosen by an experimental *S*. We would naturally expect the *S* to modify his impression of the *SP* in a negative direction, and we have seen that this is indeed the case. But the intensity or degree of negative feelings expressed depends on the implied value *S* attaches to his own eliciting behavior. The more *S* is convinced that his arguments are reasonable, just, and valid, the more righteous and therefore intense we would expect his hostility to become. In the realm of communication about controversial issues, the degree of consensual support for an argument is an important source of its perceived validity (see, for example, Festinger, 1954). The presence or absence of such social support was the main experimental variation in the present study. The results show that *S*s who perceive support for their behavior respond with greater hostility than those who perceive no such support. As predicted, this differential hostility was reflected in the *S*s' verbal impressions of the derogating *SP* and in their attempts to characterize the *SP* on a rating scale (though with the latter measure those who initially disliked the *SP* showed no such differential change).

The present results fit nicely into the discussion by Jones and Thibaut (1958) of the importance of interaction context in person perception, and the method for demonstrating these contextual effects is basically similar to that employed by Jones and Daugherty (1959) and Jones, Hester, Farina and Davis (1959). However, the obtained findings can be accommodated by a number of theoretical positions and seem not to embarrass any others. It is possible, for example, to interpret the results by a simple extension of Heider's (1958) notions of cognitive balance in the P-O-X unit. The individual *S* (*P*) is led to dislike or devalue *SP* (*O*) because they disagree about *S*'s verbal product (X). Following Cartwright and Harary (1956), this particular *P-O-X* unit can be represented as an *s*-digraph in which *P* dislikes *O*, *O* dislikes *P*, *P* likes *X*, and *O* dislikes *X*. Given the fact that *P* likes *X* (*S* values his own chosen arguments) and that *O* dislikes *P* and *X* (as revealed by *SP*'s derogation of both the arguments and the person who thought them up), *P* must dislike *O* for the structure to be balanced. In order to handle the present experimental results, however, it is necessary to introduce a further quantitative assumption: the magnitude of *P*'s liking for *X* will determine the magnitude of his *dis*like for *O*.

In the context of frustration-aggression theory, the present results might be viewed in terms of the degree to which *S*'s hostility toward *SP* is inhibited or facilitated by the Group Support manipulation. *S*s in the Group Support condition may be compared with the members of a cohesive in-group who are secure in the conviction that their beliefs and behavior are not only meaningful but morally right. A considerable amount of anecdotal evidence on witch hunts, lynchings, etc., could be cited to make the point that hostility toward out-group members can be vastly intensified when justified as serving a righteous moral passion. This could be true both because the threat to a highly moral position is viewed as more "frustrating" than the threat to a position more neutral in morality, and because the in-group provides direct positive sanctions for the expression of aggression in the service of a moral cause. The present results provide circumstantial evidence for both of these effects in the laboratory context. There is a difference between the private and the public ratings, the latter seeming to facilitate hostility in the Group Support and inhibit it in the No Support condition. However, the fact that the private ratings show the same difference to a significant

if lesser degree, indicates that the present results are not entirely the effect of the group's power to release or suppress aggression in its members. At least we may speculate that this power becomes internalized as the individual responds to others in terms of the normative support inherent in his position.

SUMMARY

The present experiment was concerned with variations in person perception as a function of the perceiver's role in instigating a stimulus person to hostility. While the stimulus person's hostile remarks were always constant across *S*s, the perceiver was led to believe that his instigating behavior either was or was not sanctioned by a group of fellow *S*s. This variation in group support was accomplished by manipulating the extent to which the other group members agreed with a set of arguments transmitted by the perceiver to the stimulus person.

The major findings were generally in line with the hypotheses in that:

1. There was a tendency for the *S*s' evaluative ratings of the stimulus person to become more negative when they perceived that they had group support for their communication (which prompted the derogation by the stimulus person) than when they perceived no such support. This difference was significant when consideration was restricted to those *S*s in each condition who were initially favorably disposed toward the stimulus person.

2. In postderogati free-response "personality sketches," Gro Support *S*s evidenced significantly greater hostility toward the stimulus person than did the No Support *S*s.

3. When anticipating the publication of their judgments to the other members of their group, the Group Support *S*s became somewhat more hostile, while the No Support *S*s became considerably less hostile.

In discussing these findings, the importance of considering the interaction context in studying person perception was stressed. The results were also discussed in terms of cognitive balance and frustration-aggression theories.

REFERENCES

CARTWRIGHT, D., & HARARY, F. Structural balance: A generalization of Heider's theory. *Psychol. Rev.*, **63**, 1956, 277–293.

FESTINGER, L. A theory of social comparison processes. *Hum. Relat.*, **7**, 1954, 117–140.

HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

JONES, E. E., & THIBAUT, J. W. Interaction goals as bases of inference in interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford: Stanford Univer. Press, 1958, 151–178.

JONES, E. E., & DAUGHERTY, B. N. Political orientation and the perceptual effects of an anticipated interaction. *J. abnorm. soc. Psychol.*, **59**, 1959, 340–349.

JONES, E. E., HESTER, S. L., FARINA, A., & DAVIS, K. E. Reactions to unfavorable personal evaluation as a function of the evaluator's perceived adjustment. *J. abnorm. soc. Psychol.*, **59**, 1959, 363–370.

(Received August 10, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 83-86

DIFFERENTIAL LEARNING AND FORGETTING AS A FUNCTION OF THE SOCIAL FRAME OF REFERENCE

R. KANUNGO AND J. P. DAS
Ravenshaw College, India

RESEARCH on the relationship between the affective tone of material to be learned and retention does not permit any definite conclusion. There have been several obstacles. First, it is very difficult to isolate experimentally the dimension of affectivity from other variables affecting the learning and forgetting processes. Conflicting results and interpretations have also arisen from ambiguities concerning the nature of the feeling tone of the materials used in experiments and the way the materials actually affect the subjects studied.

The usual procedure to test the influence of affective tone on learning and retention is to choose a number of items that have been rated as pleasant, unpleasant, and neutral. These items are learned by the subject and subsequently they are recalled, relearned, or recognized after a specified time interval. In general, the results of various experiments using their procedure go to show that retention decreases in the order of Pleasant-Unpleasant-Neutral. Other experiments using either free recall of past experience or the technique of pairing affective materials with neutral material have yielded conflicting results (Osgood, 1958, p. 571-572).

Seeking to overcome such disagreement, Edwards (cf., Underwood, 1949, p. 523) restated the problem of differential forgetting of affective material not in terms of affect, but "in terms of whether or not the experience or material learned conforms with the individual's frame of reference," (p. 523) that is, his system of evaluation including his values and attitudes.

Osgood (1958) distinguishes between two facets of the problem, one dealing with materials having felt affect and the other dealing with materials having judged affect. Edward's explanation in terms of frames of reference seems most appropriate to retention and forgetting of material having judged affect on "an intellectual or esthetic level." From this point of view, the learning and retention of material judged to have affective tone is influenced by the person's social frame of reference rather than by the feeling tone of the material itself. The present investigation is directed to explore the implications of this explanation of retention of affective materials. *Ss* were selected from two caste groups in India. The materials to be memorized consisted of favorable and unfavorable stereotypes concerning each of the caste groups, and also concerning another plausible but fictitious caste group.

The following hypotheses were formulated:

1. No significant difference should be obtained between learning and retention of favorable and unfavorable adjectives for the nonexistent caste.

2. Regardless of the judged affect, in general, stereotypes concerning one's own group should be better learned and remembered than the stereotypes concerning the other group.

3. Favorable stereotypes for one's own group and unfavorable stereotypes for the other group should be better recalled than favorable stereotypes for the other group and unfavorable stereotypes for one's own group.

4. The effects of Hypothesis 3 should be shown more clearly in delayed recall than in immediate recall.

METHOD

Subjects

Twenty-five Brahmin and 25 Karan undergraduate students served as *Ss*.[1]

Materials

A list of 20 favorable and 20 unfavorable adjectives (control list) was chosen. These were labelled as stereotypes of a fictitious caste group named "Talami." Another list (test list) consisted of two sets of stere-

[1] Brahmin, the priest's caste, and Karan, the caste of the clerks and ministers to the King in old India, are the two top classes in the part of the country in which the study was conducted. Each of the two castes considers itself as the aggressive competitor of the other in social, educational, occupational, and professional spheres of life. A certain amount of antagonism exists between them, which has been taken advantage of in the present study.

otypes, each set consisting of 20 adjectives. One set included 10 favorable and 10 unfavorable adjectives often ascribed to the Brahmin caste, and the other set included 10 favorable and 10 unfavorable adjectives similarly ascribed to the Karan caste.[2] All adjectives were in Oriya language, the mother tongue of the *Ss*. These favorable and unfavorable stereotypes relating to the two castes were taken from an earlier study.[3]

Procedure

The design of the experiment was determined by the need to provide different frames of reference to *Ss* who were engaged in learning and recalling materials that had different judged affect for them. Therefore the study had two stages. At the initial stage the *Ss* were presented the list of 40 adjectives labelled as stereotypes of the fictitious caste "Talami" with the following instructions:

> In a particular place in India there is a group of people known as Talami. Authoritative reports about them indicate that they possess the following attributes which you shall hear one by one. Please try to be attentive to what is being said about the group and try to remember them to the best of your ability. This is a test of your memory. Now listen.

The list of 40 adjectives was then presented to the *Ss* by tape recorder. Word order was randomized using a table of random numbers. After eight consecutive presentations of the list *Ss* were asked to write down as many adjectives of the list as they could. On the next day about 24 hours after the original learning *Ss* were again asked to recall the learned material. To avoid rehearsal within the intervening period care was taken not to give any hint about the delayed recall.

On the basis of their recall scores, both immediate and delayed, 15 Brahmin and 15 Karan *Ss* were chosen so as to form two equivalent groups for the next stage of the experiment.

The second stage of the investigation began two months after the initial test. This time the test list of stereotypes was given, with the following instructions:

> Earlier you were tested for your memory ability. To know whether, in the meantime, you have deteriorated or improved in your memory ability, we shall have a fresh test. Last time the list consisted of stereotypes towards the Talami group whom you never knew before. But this time you will be given two sets of stereotypes, one relating to Brahmins and another relating to Karans whom you know quite well. Try to remember what is said about the two castes separately. As this time we just want to check our earlier results, you would not be given a delayed recall test.

[2] The lists of stereotypes comprising the initial and final tests have been deposited with the American Documentation Institute. Order Document No. 6316, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm, or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[3] Unpublished research on social tension, by R. Rath (1958).

TABLE 1

MATCHING OF GROUPS ON RECALL SCORES, INITIAL TEST

(In each group, $N = 15$)

Recall	Caste	*M*	*SD*	*t*	*P*
Immediate	Karan	25.66	3.52	.01	NS
	Brahmin	25.62	4.82		
Delayed	Karan	22.13	3.68	.28	NS
	Brahmin	21.66	4.75		

TABLE 2

RECALL SCORES OF BRAHMINS AND KARANS

(In each group, $N = 15$)

Caste	Nature of Stereotypes	Immediate Recall *M*	Immediate Recall *SD*	Delayed Recall *M*	Delayed Recall *SD*
Brahmin	Favorable stereotypes towards themselves	5.33	1.63	3.67	.45
	Unfavorable stereotypes towards themselves	5.27	1.97	4.73	2.13
	Favorable stereotypes towards others	3.6	2.3	3.13	2.24
	Unfavorable stereotypes towards others	4.67	1.35	4.07	1.77
Karan	Favorable stereotypes towards themselves	5.8	1.55	4.40	1.86
	Unfavorable stereotypes towards themselves	5.93	1.70	3.67	1.48
	Favorable stereotypes towards others	4.6	1.96	3.27	1.65
	Unfavorable stereotypes towards others	6.2	1.30	5.53	1.51

Then the two sets of stereotypes of the test list, one relating to Brahmins and the other to Karans, were presented by tape recorder. Each set contained equal number of favorable and unfavorable stereotypes in randomized order. The two sets were presented in A-B-B-A order: stereotypes relating to Brahmins, then stereotypes relating to Karans, stereotypes relating to Karans, stereotypes relating to Brahmins. This sequence was repeated consecutively for four times. The instructions were framed to minimize rehearsal that might have occurred had the *Ss* anticipated measurement of delayed recall. After about 24 hr., however, measures of delayed recall were taken.

RESULTS AND DISCUSSION

An examination of Table 1 reveals that an attempt to select equated groups in the Initial test was successful.

The first hypothesis predicted no significant difference in learning and recall of favorable

as against the unfavorable adjectives ascribed to the fictitious caste. This was borne out by a t test.

The results of the main experiment are presented in Table 2. An analysis of variance was performed on the scores of the main experiment and the results are given in Table 3. A 2^n factorial design was adopted (Edwards, 1950) since there were four main variables each dichotomously divisible. All the main effects and only one interaction were significant. The second hypothesis, which predicted a significant difference between recalling adjectives relating to one's own caste vs. those relating to the other caste, is supported by the significant F for main Effect 4. Each group appears to be more interested in what is being said about itself, irrespective of whether the adjectives are favorable or unfavorable.

The only significant interaction may be taken to support the third hypothesis which predicted a differential recall of favorable and unfavorable adjectives relating to ones own and to the other group. But does this overall significant interaction mean that for each caste and both for immediate and delayed recall, favorable adjectives for one's own group and unfavorable ones for the other group are better recalled? Further analysis presented in Table 4 shows that this is not the case. Contrary to predictions, there was no difference in the immediate recall of unfavorable stereotypes relating to one's own and to the other group. Only in the Karan caste was the predicted difference in favorable stereotypes obtained for both immediate and delayed

TABLE 3
ANALYSIS OF VARIANCE OF RECALL SCORES

Source	Sum of Squares	df
Immediate & delayed recall	74.8167***	1
Castes	22.8167**	1
Favorable & unfavorable	36.8167***	1
Favorable and unfavorable towards themselves and favorable and unfavorable towards others	13.0667*	1
1 × 2	5.400	1
1 × 3	0.600	1
1 × 4	7.3500	1
2 × 3	0.0667	1
2 × 4	10.4167	1
3 × 4	28.0167**	1
1 × 2 × 3	1.3500	1
1 × 2 × 4	0.2677	1
1 × 3 × 4	0.0667	1
2 × 3 × 4	11.2667	1
1 × 2 × 3 × 4	7.3500	1
Residual	720.054	224
Total	939.734	239
Residual msv.	3.2145	

Note.—F significant:
* $P < .05$
** $P < .01$
*** $P < .001$

TABLE 4
RECALL OF FAVORABLE AND UNFAVORABLE STEREOTYPES AS DIRECTED TOWARDS OWN CASTE OR TOWARDS OTHER CASTE

Caste	Nature of Stereotypes	Immediate Recall					Delayed Recall				
		Mean$_D$	SD_D	SE_M	df	t	Mean$_D$	SD_D	SE_M	df	t
Karan (N = 15)	Favorable stereotypes towards themselves > towards others	1.2	1.82	0.47	14	2.35*	1.13	1.88	0.48	14	2.35*
	Unfavorable stereotypes towards others > towards themselves	0.26	1.45	0.37	14	0.70	1.86	0.99	0.25	14	7.44***
Brahmin (N = 15)	Favorable stereotypes towards themselves > towards others	1.73	1.43	0.37	14	4.67***	0.53	1.27	0.32	14	1.65
	Unfavorable stereotypes towards themselves > towards others	0.60	1.82	0.44	14	1.36	0.66	1.23	0.31	14	2.12

* $P < .05$.
*** $P < .001$.

TABLE 5

RATING OF STEREOTYPES BY 29 JUDGES INTO FIVE CATEGORIES

Frequency of Use	No. of Ratings Favorable Stereotype ($N = 20$)		No. of Ratings Unfavorable Stereotype ($N = 20$)	
Very frequently used	122	230	249	378
Frequently used	108		129	
Neither frequently nor infrequently used	119		75	
Less frequently used	117	231	88	127
Least frequently used	114		39	
Total	580		580	

recall. However, except for the recall of unfavorable stereotypes by Brahmin *S*s, the differences were in the predicted direction.

The fourth hypothesis is only partially upheld by the significant difference in the delayed recall of unfavorable stereotypes by the Karan *S*s.

Besides confirming the hypotheses under study, the results of the analysis of variance pointed up two significant main effects which were not predicted: castes and favorable-unfavorable stereotypes. Although the castes were matched in the initial test, the Brahmin *S*s are found to have a significantly lower recall score than the Karan *S*s in the main experiment. Perhaps the Brahmin group is more heterogeneous, for on the whole, it has larger *SD*s than the Karan group (see Tables 1 and 2).

The significant difference in the recall of favorable and unfavorable stereotypes calls for closer attention. The most plausible explanation as to why unfavorable stereotypes are better recalled may be that *S*s are more familiar with the unfavorable stereotypes, possibly because these are more often used in spoken language. In order to test this hypothesis, all the 40 stereotypes used in the main experiment were given to 29 judges in a predetermined random order for rating. The judges, who were college teachers, rated each of the sterotypes on a five-point scale representing the extent of use of these words in common spoken language. The results are presented in Table 5 and can be seen to support the hypothesis that unfavorable stereotypes are more familiar. Leaving out the middle category, the unfavorable stereotypes receive 378 ratings as frequent as against 127 infrequent ratings; but the comparable figures for favorable stereotypes are 230 and 231. Again, the unfavorable stereotypes get 249 ratings as "very frequently used" whereas the corresponding figure for favorable stereotypes is 122. It was indeed a surprising revelation to the authors that in their language unfavorable stereotypes like traitor, swindler, and bribe-taker find greater favor than such favorable ones like cultured, hospitable, or artistic.

SUMMARY

The present study was aimed at demonstrating how learning and recall become selective because of the relevant frames of reference. Two groups of 15 *S*s each, were selected from two competing caste groups, Brahmins and Karans, and were required to learn and recall 20 favorable and 20 unfavorable familiar stereotypes concerning themselves. The groups were matched on the basis of their recall scores in a preliminary test.

It was hypothesized that (*a*) there should be no difference in learning and retention of favorable and unfavorable stereotypes for a fictitious caste, (*b*) stereotypes relating to one's own caste should be better learned and remembered than those relating to the other caste, (*c*) favorable stereotypes for one's own caste and unfavorable stereotypes for the other caste should be better learned and recalled than favorable stereotypes for the other caste and unfavorable stereotypes for ones own caste, and (*d*) this will be shown more clearly in delayed recall than in immediate recall.

A 2^n factorial analysis of variance showed that Hypotheses 1, 2, and 3, were valid, the last one only generally. Hypothesis 4 was only partially supported.

Unfavorable stereotypes were better recalled than the favorable ones. A supplementary test suggested that this was so because the unfavorable stereotypes were more often used

in the language of the *S*s than the favorable ones.

REFERENCES

EDWARDS, A. L. *Experimental designs in psychological research.* New York: Rinehart, 1950.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1958.

UNDERWOOD, B. J. *Experimental psychology.* New York: Appleton-Century-Crofts, 1949.

(Received July 21, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 87-92

INTERPERSONAL KNOWLEDGE AND RATED LEADER POTENTIAL[1]

MORRIS SHOWEL

Human Resources Research Office, George Washington University[2]

IT IS a widely accepted principle of military leadership that a leader should know his men. In psychological terms, this could be translated to mean that a leader should demonstrate "empathy," "social sensitivity," or "social perception." In spite of a considerable amount of research on social perception and the relation between ability in this area to some criterion of effectiveness (Taft, 1955), the validity of the principle has not yet been satisfactorily demonstrated. Moreover, considerable criticism has been levied against the conceptual approaches and the research procedures used to study the problem in the past. (Cronbach, 1955; Gage & Cronbach, 1955.)

This paper presents a somewhat different approach toward measuring social perception and relating it to a criterion of effectiveness. First, the concept of social perception has been broadened to include not only perception of attitudes, opinions, and preferences, but of demographic data as well. Second, rather than dealing with accuracy of social perception of group characteristics, this study deals with accuracy of social perception about the particular individuals comprising the group. Bush (1954) used a similar procedure and found a positive correlation between teacher effectiveness and the amount of knowledge teachers had about their students. Using the procedure just described seems to avoid some of the criticisms raised by Gage and Cronbach (1955) to previous research in this area: the problem of real similarity, of stereotype accuracy, and of securing an independent criterion of effectiveness in interpersonal relations.

[1] The opinions and conclusions contained herein are those of the author and are not to be construed as representing those of George Washington University or the Department of the Army.

[2] Located at United States Army Leadership Human Research Unit, P.O. Box 787, Presidio of Monterey, California.

METHOD

Subjects and Setting

Subjects for this study were infantry trainees in the process of completing their six-month tour of duty. These trainees comprised two of the five platoons of one training company. Each of the platoons contained four squads, and each squad contained approximately ten nonleader trainees. The company was led by a company commander (officer), whose chief administrative assistant was a first sergeant (noncommissioned officer). Each platoon was led by a platoon leader (officer) who was assisted by a cadre platoon sergeant (noncommissioned officer). In addition, each platoon had a trainee platoon sergeant and a trainee platoon guide who acted in a leadership capacity below the platoon leader and the cadre platoon sergeant. Each squad was led by a trainee squad leader.

Platoons were housed in separate barracks, two squads per floor. Squad leaders lived with their men. The trainee platoon sergeant and the trainee platoon guide lived in separate rooms in their platoon's barracks. The cadre platoon sergeant, platoon leader, first sergeant, and company commander did not live in the barracks.

All of these personnel had served together for a minimum of six weeks, some of the trainees for as long as six months. The intimate nature of barracks living and the extended period of time during which trainees served together provided considerable opportunity for the trainees to get to know one another.

Interpersonal Knowledge Questionnaire

Accuracy of social perception was measured by means of an Interpersonal Knowledge Questionnaire. The information asked for in the questionnaire was selected on the basis of three criteria: (*a*) the information would be useful for a leader to know; (*b*) it could easily be verified; (*c*) it could be secured via a questionnaire or brief interview. The questionnaire included a report on self, on the men in the squad, on the trainee platoon sergeant and guide, and on the cadre (cadre platoon sergeant, platoon leader, first sergeant, and company commander).

Report on self. All trainees answered the following questions about themselves: (*a*) first name; (*b*) last name; (*c*) whether or not during the past week they had been on KP; (*d*) whether or not during the past week they had been on sick call or in the hospital; (*e*) whether or not during the past week they had been on pass or leave; (*f*) their rifle qualification score; (*g*) the amount of schooling they had completed; (*h*) their main job or activity before entering the Army; (*i*) their

principal hobby or interest; (*j*)their ambition for a future civilian career.

Report on trainee squad leaders and nonleaders. Trainee squadleaders and trainee nonleaders were given the last names of the men in their squad, including the squad leader, and then asked to answer all of the other questions about these people. Reports on trainee squad leaders and nonleaders were then compared with the self-reports to determine the amount of interpersonal knowledge possessed. As a result of the relatively objective nature of the data, all scoring was done by one person. Items #8, #9, and #10 required the most interpretation on the part of the scorer. The objective nature of the data also makes it evident that relatively little extrapolation was required of the *S* when reporting on others. Each correct answer was awarded one point. Neither blanks nor incorrect answers were counted. Since the squads differed somewhat in size, a correction for these differences was handled by summating the points received by each man, dividing this total by the number of men in his squad, and then multiplying it by 10, to give a score with a ceiling of 90 points.

Report on trainee sergeants and guides. Trainee squad leaders and trainee nonleaders were asked to answer Questions 1, 2, and 4 through 10 about their trainee platoon sergeant and trainee platoon guide. (Trainee guides and sergeants do not serve as KPs.) Scoring was done in the same manner, for a possible 18 points.

Report on cadre. Trainee squad leaders and trainee nonleaders were asked to write: (*a*) the first name; (*b*) the last name; (*c*) the exact rank of their company commander, first sergeant, platoon leader and cadre platoon sergeant.[3]

Reports on cadre were then compared with information secured from these people via interview. Scoring was done in the same manner, for a possible 12 points.

Total knowledge. A total knowledge score was computed by summing the scores received by each *S* on cadre, guide and sergeant, and squad leaders, and nonleaders (corrected for size of squad). Knowledge about trainee squad leaders and nonleaders contributed most heavily to the score.

Rated Leader Potential

Five measures of leader potential were secured.

Trainee squad leaders and nonleaders ratings. Each member of a squad, including the squad leader, ranked the men in his squad, including the squad leader, in terms of the kind of a squad leader the man would make. (At the time of the study, the squad leaders were actually *acting* squad leaders.) Following the ranking, the *S*s assigned each man whom he had ranked a numerical weight. These weights ranged from 1 ("the very best type of squad leader") to 7 ("the worst kind of squad leader"). An individual's trainee squad leader and nonleader rating was the mean of the weights he had been given by his squad mates. His rank order, as such, was not used in the subsequent analysis.

[3] Company commanders, first sergeant, platoon leader, and platoon sergeant are job titles, not rank designations.

Trainee sergeants and guides ratings. The trainee platoon sergeant and the trainee platoon guide ranked and rated the trainees in their platoon in the same manner. An individual's trainee sergeants and guides rating was the mean of these two ratings.

Cadre rating. The cadre platoon sergeant and the platoon leader ranked and rated the men in their platoon and cadre ratings were computed in the same manner as for trainee sergeants and guides.

Total superiors' ratings. A total superiors' rating was computed by summing the weights assigned by trainee sergeants, trainee guides, platoon sergeant, and platoon leader, and dividing this value by four.

Leader reaction test. Forty-three of the trainees were tested by means of a leader reaction battery (United States Department of the Army, 1958a). Three of the tests in the battery required the *S* to perform as a leader in tactical situations, two of the tests required him to take part in leaderless discussion groups, and one test required him to lead a squad through dismounted drill. The followers in the tactical tests (three different persons for each test) were not acquainted with the *S*. The followers in the two discussion groups and in the dismounted drill test were generally from the same platoon, but not necessarily from the same squad. Each of the six tests lasted approximately 15 min. Ratings on all six tests were made by trained raters who had never seen the *S* prior to the test. A *S*'s leader reaction test score is the weighted average of his scores on all six tests.

Intelligence

In order to control for intelligence, the general-technical aptitude area score of each *S* was secured. The GT score is made up of the verbal intelligence and arithmetic reasoning examination scores received by each trainee when he entered the Army.

Method of Analysis

Because of the prominent and highly visible position occupied by trainee platoon sergeants, platoon guides, and squad leaders, it was decided to remove them from the larger trainee nonleader group before examining the relationship between the knowledge variables, the leader potential variables, and intelligence by means of product-moment and partial correlation coefficients.

Comparisons were also made between the average scores of trainee squad leaders and trainee nonleaders. Significance of differences between means, when the variances and *N*s are unequal, was tested by means of a procedure proposed by Cochran and Cox and cited by Johnson (1949).

RESULTS AND DISCUSSION

Leader-Nonleader Comparisons

As can be seen in Table 1, the average scores of the trainee squad leaders were consistently more favorable than the average scores of the nonleader trainees. It appears that trainee infantrymen acting in the capacity

TABLE 1

SCORES RECEIVED BY TRAINEE SQUAD LEADERS AND NONLEADERS ON INTERPERSONAL KNOWLEDGE QUESTIONNAIRE, GT, AND LEADER POTENTIAL

	Trainee						t	p
	Squad Leaders			Nonleaders				
	M	*N*	*SD*	*M*	*N*	*SD*		
Knowledge Regarding								
1. Cadre–12 points possible	8.8	7	1.46	7.8	78	2.25	2.38	*NS*
2. Trainee sergeants and guides–18 points possible	12.8	7	2.36	8.1	78	3.63	4.52	.01
3. Trainee squad leaders and/or nonleaders–90 points possible	47.0[a]	7	7.88	37.1	78	14.46	2.73	.05
4. Total knowledge–120 points possible	68.7	7	8.90	53.0	78	18.27	3.45	.01
Intelligence								
5. GT examination score	111.2	8	19.86	97.8	78	22.36	1.69	*NS*
Leadership ratings given by								
6. Trainee squad leaders and/or nonleaders (possible range 1 to 7)[b]	2.1[a]	8	.82	3.8	78	.97	5.18	.001
7. Trainee sergeants and guides (possible range 1 to 7)[b]	2.2	8	.83	4.5	78	1.32	6.63	.001
8. Cadre (possible range 1 to 7)[b]	2.5	8	.61	3.9	78	.77	5.64	.001
9. Total leaders (possible range 1 to 7)[b]	2.3	8	.62	4.2	78	.94	7.34	.001
10. Leader reaction test	105.0	4	10.82	76.1	43	16.01	4.29	.02

[a] Trainee squad leaders answered interpersonal questions about trainee nonleaders only and only received ratings from trainee nonleaders.

[b] Lower scores indicate higher ratings.

of squad leader receive higher leader potential ratings, perform better in the leader reaction tests, and have more accurate interpersonal knowledge about their associates than trainee infantrymen not acting in the capacity of squad leader. There is no significant difference between the two groups with regards to intelligence.

At least three explanations are possible for these findings. First, possession of interpersonal knowledge could increase the trainee's ability to deal with others, and lead the trainee's superiors to consider him a potential leader and place him in a leadership position. Second, it is possible that the increased amount of interpersonal knowledge was secured *after* the trainee is in a leadership position. The need to deal with others, both subordinates and superiors, may lead the trainee to acquire the knowledge necessary to do the job. On the other hand, the higher leader potential ratings could be a function of being in an official leadership position. The position of squad leader is conspicuous, and a man could receive a high rating by virtue of occupying and being highly visible in this leadership position. A third possible explanation might be that ratings, position, and knowledge are all symptomatic of another underlying variable.

Nonleader Relationships

In order to control for the highly visible position occupied by the trainee leaders, the remaining discussion is confined to data (Table 2) involving trainees who did not hold a leadership position at the time the study was conducted.

Knowledge. The three independent measures of interpersonal knowledge (Variables 1, 2, and 3) are positively and significantly correlated with one another. Nonleaders who were well informed about the cadre were also well informed about the trainee sergeants and guides, and about the trainee squad leaders and nonleaders. The total knowledge score (Variable 4) shows a moderately strong relation to Variables 1 and 2, and a near perfect relation to Variable 3. It should be remembered that the correlations of the three knowledge scores (Variables 1, 2, and 3) with the total knowledge score (Variable 4) are part-whole correlations and hence spuriously high.

TABLE 2
CORRELATIONS BETWEEN INTERPERSONAL KNOWLEDGE, LEADER POTENTIAL, AND INTELLIGENCE FOR NONLEADER TRAINEES

	1*	2*	3*	4*	5*	6*	7*	8*	9*	10**
Knowledge Concerning										
1. Cadre		.46	.58	.67	.40	.18	.14	.16	.15	.34
2. Trainee sergeants and guides			.58	.71	.44	.21	.31	.15	.27	.25
3. Trainee squad leaders and nonleaders				.98	.50	.32	.43	.15	.36	.37
4. Total knowledge					.53	.32	.42	.17	.36	.38
Intelligence										
5. GT examination score						.33	.27	.21	.28	.44
Leadership ratings given by										
6. Trainee squad leaders and nonleaders[a]							.71	.51	.72	.07
7. Trainee sergeants and guides[a]								.52	.92	.14
8. Cadre[a]									.80	.13
9. Total leaders[a]										.15
10. Leader reaction test										

* *N*-78, .233 significant at .05 level; .320 significant at .01 level.
** *N*-43, .304 significant at .05 level; .393 significant at .01 level.
[a] Negative coefficients involving Variables 6, 7, 8, and 9 have been reversed.

Leader ratings. The three ratings of leader potential (Variables 6, 7, and 8), are positively and significantly correlated with one another. Nonleader trainees rated as having high leader potential by the cadre are also rated as having high leader potential by trainee sergeants and guides and by trainee squad leaders and nonleaders. Some differences might have been expected in view of the background differences between trainees and cadre. The trainees were all six-month soldiers, having enlisted for a six-month tour of duty in preference to being drafted for two years or enlisting for three or more years. The cadre, on the other hand, were more professional military men. Nevertheless, it appears that cadre and trainees used similar standards in evaluating the leader potential of the trainees.

While trainees and cadre appear to look at leadership in a somewhat similar manner, it is nevertheless apparent that the two trainee groups (sergeants and guides, and squad leaders and nonleaders) are more in agreement with one another than are either with the cadre. The difference in background between trainees and cadre appears to be the most reasonable explanation of this pattern.

Leader reaction test. Scores on the leader reaction battery (Variable 10) do not correlate significantly with any of the other ratings of leader potential. This finding suggests that the dimensions of leadership considered by the raters and those measured by the tests are quite different. It is not possible to say with any certainty what dimensions were considered by the raters when they made their ratings. It seems likely however, that the trainee raters took into consideration human relations skills when they made their ratings,[4] since they had lived and worked together for a minimum of six weeks. The performance tests, on the other hand, made few if any demands on human relations skills. The tests were primarily concerned with the trainee's ability to use instantaneously the technical military skills he had been taught. The followers in most cases were men with whom the trainee had not developed any working relationship, and the brief duration of the tests prevented him from developing such a relationship.

Knowledge and leadership. The more knowl-

[4] Another questionnaire (completed by a larger group of *Ss*) required a ranking in order of importance of 10 qualities that a squad leader should have. Both cadre and trainees placed the same five qualities in the upper half of the ranking. The most important quality for both groups was "willingness to accept responsibility," but the cadre gave it a mean rank of 2.3 while the trainees gave it a mean rank of 3.4. The second most important quality for the cadre was "to be efficient in his job" (mean rank 3.1). This quality was fifth in importance for trainees (mean rank 4.9). The second most important quality for the trainees was "to understand the men" (mean rank 4.4). This quality was fifth in importance for the cadre (mean rank 5.3). This analysis was carried out by C. W. Peterson.

edge a nonleader trainee had about his trainee squad leader and nonleader colleagues, the higher was the leader potential rating he received from *all* trainees. This relationship cannot be accounted for by the position that these nonleaders occupied, for their position in the squad and platoon, both with regard to visibility and authority, was no different from that of nonleaders with low ratings and low levels of interpersonal knowledge. Again, a number of explanations are possible. First, in the process of securing interpersonal information, the high rate might have given indication of a high degree of personal interest in their co-workers, who in turn may therefore have been more likely to rate them higher in leader potential. Second, having such information might enable a high ratee to deal more effectively with his co-workers. This in turn might lead co-workers to give these trainees higher leader potential ratings. Third, general intelligence, or some other factor, might account for both the high level of interpersonal knowledge and the high leader potential ratings.

Knowledge about trainee sergeants and guides correlated positively and significantly with the leadership ratings given by these people. Knowledge about the cadre was not significantly correlated with the leader potential ratings given by either trainees or the cadre.

Having accurate information about co-workers, either trainee or cadre, appears to have little if any bearing upon the leader potential ratings received from the cadre. It appears here that the dimension of leadership tapped by the Interpersonal Knowledge Questionnaire was not considered by the cadre when they made their ratings.

Intelligence. Intelligence, as measured by GT, correlated positively and significantly with all four measures of interpersonal knowledge, and with all but one of the five measures of leader potential. Using the total knowledge score as the measure of interpersonal knowledge, the relation between interpersonal knowledge and the various measures of leader potential, holding intelligence constant, are as follows: with trainee squad leaders and nonleaders ratings, $r = .18$; with trainee sergeants and guides ratings, $r = .34$; with cadre ratings, $r = .07$; and with the leader reaction battery scores, $r = .20$. Only the correlation with trainee sergeants and guides ratings is statistically significant.

It thus appears that a considerable portion of the variation in leader potential originally attributable to interpersonal knowledge may be attributed to intelligence. Yet not all can be so attributed. A moderate relation remains between total interpersonal knowledge (primarily a measure of knowledge about trainee squad leaders and nonleaders) and ratings given by trainee sergeants and guides. This is particularly interesting, for in many respects, the trainee sergeants and guides may have been the most valid raters. They had far greater opportunity than the cadre to observe the day-to-day performance of the nonleader trainees, and their conception of what constitutes the military requirements of leadership were probably more accurate than the conceptions of military leadership held by trainee squad leaders and nonleaders.

The obtained partial correlation of .34 between interpersonal knowledge and trainee sergeants and guides ratings is not particularly high in a practical sense, but the fact that it is significant suggests that there is some empirical support for the generally accepted leadership principle "know your men" (USA, 1958a).

SUMMARY

This study investigated the relationship between the possession of interpersonal knowledge about others and the ratings received as to leader potential. Four measures of interpersonal knowledge and five measures of leader potential were secured. *Ss* were two platoons of soldiers completing a six-month tour of duty in the Army.

The data indicate that the more interpersonal knowledge nonleader trainees had, the higher was the leader potential ratings they received from trainee leaders and trainee nonleaders. The correlation drops when intelligence is controlled, but that between total knowledge possessed and ratings received from one group of trainee leaders, trainee sergeants and guides, still remains statistically significant. It is hypothesized that the ratings given by trainee sergeants and guides are more valid than the ratings given by the cadre, by trainee squad leaders, and by trainee nonleaders.

Additional findings were that trainee squad

leaders had more interpersonal knowledge and received higher leader potential ratings than trainee nonleaders. Trainee squad leaders and trainee nonleaders did not differ significantly in regard to intelligence.

REFERENCES

BUSH, R. N. *The teacher-pupil relationship*. New York: Prentice-Hall, 1954.

CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177–193.

GAGE, N. L., & CRONBACH, L. J. Conceptual and methodological problems in interpersonal perception. *Psychol. Rev.*, 1955, **62**, 411–422.

JOHNSON, P. O. *Statistical methods in research*. New York: Prentice-Hall, 1949.

TAFT, R. The ability to judge people. *Psychol. Bull.*, 1955, **52**, 1–23.

UNITED STATES DEPARTMENT OF THE ARMY. *Military leadership*, FM 22-100. Washington, D. C.: USA, 1958. (b)

UNITED STATES DEPARTMENT OF THE ARMY, Adjutant General's Office. *Manual for administering and scoring the Leadership Performance Battery*, PT 3579. Washington, D. C.: USA, 1958. (a)

(Received July 20, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 93–101

COERCIVE POWER AND FORCES AFFECTING CONFORMITY[1]

JOHN R. P. FRENCH, JR., H. WILLIAM MORRISON,
University of Michigan

AND GEORGE LEVINGER
Bryn Mawr College

WHAT determines the conformity of a subordinate to the orders of his superior? In a more general theory, based on Lewin's constructs, we have distinguished five bases of power that might lead to such conformity: legitimate, coercive, reward, referent, and expert power (French & Raven, 1959). The theory focuses on *P*, the person being influenced by the other person, *O*. Two previous experiments demonstrated that the overt behavior of *P* conforms to the orders of *O* when this superior has legitimate power over *P* and when he has coercive power. (Raven & French, 1958a; 1958b). But there was a tendency for *P*'s private opinions to conform less when *O* used coercion. To account for these findings it seemed necessary to assume that *P*'s behavior was determined both by induced punishment forces and by opposing resisting forces. However, no measures of these intervening forces were obtained, so there was no direct test of the interpretation. The purpose of this experiment is to test directly this theory about the effects of punishment forces and resisting forces on conformity.

The structure of the theory is outlined in Fig. 1, with the independent (experimental) variables at the top of the figure, the dependent variable at the bottom, and the intervening variables in between. The solid arrows indicate the major hypotheses or assumptions, while the dotted arrows indicate hypotheses that may be derived from two other hypotheses. The formal statement of these hypotheses is given in connection with presenting the relevant results, and the numbering of the arrows in Fig. 1 corresponds to the numbers of these hypotheses. Thus, the data for each relation can be found in the section on results under the hypothesis whose number is given in parentheses. The other numbers in Fig. 1 show the derivations of the derived hypotheses.

[1] This experiment was conducted at the Research Center for Group Dynamics under contract with the Office of Naval Research.

The central part of the theory states that the amount of conformity of a subordinate to his superior depends on the balance of the punishment force and the resisting force. The punishment force is defined as a tendency toward conformity that is set up by *O*'s act of coercion—either a punishment or a threat of future punishment. A punishment is defined as an intentional act of *O* which places *P* in a region of negative valence; an accidental act of *O* which harms *P* is not perceived as a punishment. *Threats* of future punishment similarly place *P* in a region of negative valence, but only to the extent that the threatened punishment is probable; if *P* does not believe the threat, he does not anticipate punishment. If *P* perceives that *O* will probably carry out his intentional threat of punishment for nonconformity, then *P*'s conforming behavior is instrumental to avoiding the punishment. Thus, the threat sets up punishment forces that produce conformity (Hypothesis 1 in Fig. 1). But actual punishment should be effective in producing subsequent conformity only if *P* expects that the punishment will be continued or repeated (Hypothesis 2).

The strength of the *punishment force* based on an expectation of punishment by *O* and directed toward compliance with *O*'s influence attempts could not be measured directly in this experiment. Instead we calculated its strength on the assumption (cf., Fig. 1) that it depended on the valence of the punishment and the differential probability of receiving it contingent upon nonconformity vs. conformity. The formula was:

Punishment force = (valence of punishment) × (probability of punishment if nonconformity − probability of punishment if conformity)

From the assumption expressed in this formula and from Hypothesis 1 we may derive the prediction that threatened *S*s in the coercive experimental treatment will show

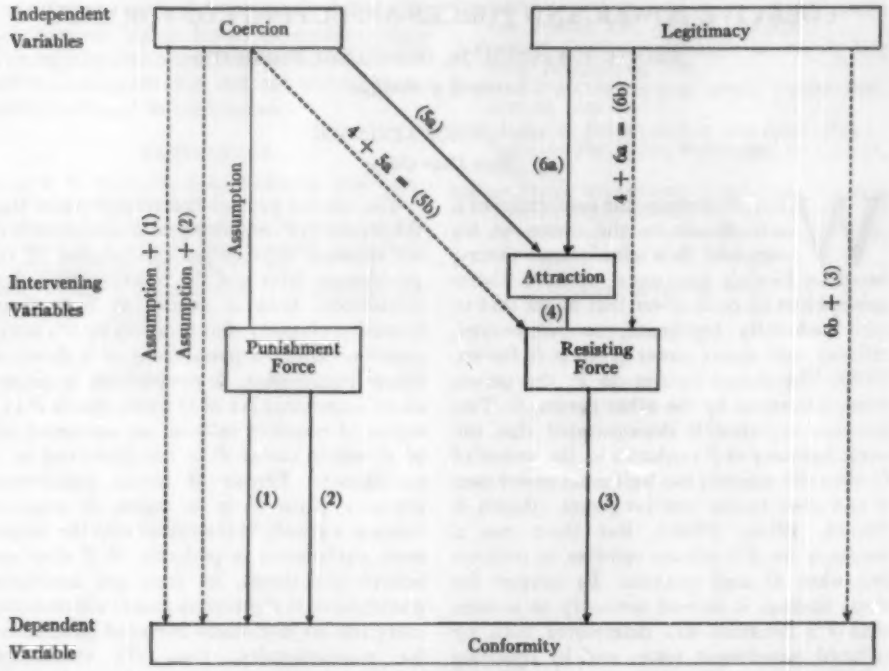

FIG. 1 Experimental variables and hypotheses testing their relationship. (The main hypotheses are represented as solid arrows with their encircled numbers attached; derived hypotheses and supplemental tests are represented as dotted arrows which are labelled so as to show the derivation, with the encircled number indicating where the test can be found in the text.)

more conformity than *S*s in a noncoercive treatment. Similarly, the assumption and Hypothesis 2 leads to the derivation that no such difference will occur after the punishment is over and no further punishment is possible.

A *resisting force* is defined as a force resulting from the same act of *O* that gave rise to the punishment force but in a direction opposite to that of the punishment force. Such resisting forces should reduce conformity (Hypothesis 3).

But what determines the strength of resisting forces? In the previous experiment low attraction to the supervisor was produced by both the supervisor's utilization of coercion (Hypothesis 5a) and by his illegitimate occupancy of the supervisory role (Hypothesis 6a). Therefore, it seemed reasonable to hypothesize that low (or negative) attraction produced resistance (Hypothesis 4) and that attraction mediates the effects of coercion on resisting force (Hypothesis 5b) and the effects of legitimacy on resisting force (Hypothesis 6b). However, it may be that *P*'s values are more important in mediating the effects of legitimacy on resisting forces. Perhaps the more *O*'s coercion is perceived as legitimate or "right," the more would resistance to his coercive influence attempts be incompatible with *P*'s value system regarding *O*'s role and the appropriate use of punishment to gain conformity.[3]

METHOD

Design and Manipulations

Ninety male undergraduate *S*s met in groups of 6 to 10 persons and worked at sorting IBM cards by hand. *S*s had been obtained through volunteering, but

[3] An alternative interpretation, formally similar to the derivation of resisting forces presented here, is that coercion results in the reduction of other kinds of power. If personal attraction is another basis for power (French & Raven, 1959), the use of punishment might lead to a decrease in attraction and consequently to a decrease in conformity.

TABLE 1
NUMBER OF *S*s IN EACH EXPERIMENTAL CONDITION

Sequence of Manipulations	50¢ Fine Is Legitimate	25¢ Fine Is Legitimate
Evaluation too slow, threat, evaluation too slow, 50¢ fine	33 (Legitimate Fine)	22 (Illegitimate Fine)
Evaluation too slow, no threat, evaluation too slow, no fine	11 (No Fine)	10 (No Fine)
No evaluation, no threat, no fine	10 (Control)	4 (Control)

when scheduled for their appointments they were told they would receive $1.00 for working on an actual project at the Institute for Social Research.

The coercive power of the supervisor (a collaborator of *E*) was set up by telling *S*s that he could legitimately fine them some part of their pay if they did not perform their job satisfactorily. The amount of the supervisor's *legitimate coercive power* was varied in *E*'s initial statement to *S*s. Some of the groups were told that the supervisor was allowed to fine them up to 50¢, while others were told that the maximum fine was 25¢. Thus, the supervisor's subsequent 50¢ fine was legitimate in the former groups but illegitimate in the latter. The second independent variable in the 2 × 3 design was the supervisor's use of *coercive power*, as shown in Table 1. For all *S*s the intervening variables—*punishment force, attraction to the supervisor* and *resisting forces*—were measured by a questionnaire after the completion of the task. Finally, the dependent variables of *conformity* were measured from *S*s' production records and from their own evaluations of their production on questionnaires between work periods.

Experimental Instructions

E's *introduction.* *E* first presented a cover story intended to give a plausible reason for having asked *S*s to come, yet to draw their attention away from the critical manipulations. *E* said that he was studying "different aspects of production in the performance of routine tasks" including the degree of communication among workers, the length of rest periods, and the nature of time orientation. Accordingly, "in this group" there would be no communication among workers, long rest periods, and a metronome beating during the work session.

Frank (1944) has shown that *E* has a wide domain of legitimate power that prevents resistance. In order to give *S*s the freedom to resist the supervisor's influence attempts without feeling they were sabotaging *E*'s study, *E* made it clear that his interest in their work was separate from the supervisor's. *E* would be interested only in the effects of the above three factors, and *S*s' work for the supervisor was to be "entirely between you and him."

It was necessary that the extent of the supervisor's legitimate coercive power be set up by an external source. For this reason, *E* explained that since the supervisor "is so late today" he would go ahead and tell *S*s about their pay. They were to get $1.00, but since he had complained about some *S*s' poor work the supervisor had demanded the right to fine for unsatisfactory performance. The supervisor's project head and *E* had together decided that although the *S*s were entitled to their pay, the supervisor should have the right to fine not more than 50¢ (or 25¢, depending on the experimental condition) of an *S*'s pay. Any more than that would certainly be unfair.

E then repeated his instructions about no talking, filling out forms during rest periods, etc. At this time the supervisor entered, apologized for being late, and was introduced to the workers. *E* then left.

Supervisor's directions. The supervisor let it be known that he was not concerned with *E*'s experimental conditions of communication, rest, and distraction; he cared only about getting the sorting job done. He explained that IBM cards had to be hand-sorted in order to separate them by number of punches. He then went on to show *S*s how they were to sort the cards, admitting that the task would be tedious, but insisting that the cards should be sorted correctly and not damaged. *S*s were to work *fast* and *carefully*, but not to worry about individual mistakes since there would be some checking of their work later. Without mentioning any amount of possible fine, the supervisor mentioned that if any fine for poor work would be necessary he would warn the individual concerned beforehand.

Work arrangements. *S*s were then seated at a table separated from each other by boothlike partitions, so that they could neither see each other's work nor conveniently speak to each other. During rest periods *S*s remained seated at their places reading magazines. To abide by the rules of no communication, the supervisor passed out all his evaluations, threats, and information about punishments in written notes. Each *S*'s pay was placed in a small container in his work booth so that the supervisor could administer the fines by actually taking away 50¢.

Schedule of Work Sessions

The schedule of work and rest was as follows:

Work Period 1—6 min.

Rest Period 1—5 min., first questionnaire

Work Period 2—6 min. [base period for measuring change in production following threat]

Rest Period 2—5 min.; No Evaluation *S*s received no communication. No Fine *S*s received evaluation form, "You're working *too slowly*. You won't be fined, but *go faster*." Fine *S*s received evaluation form, "You're working *too slowly*. *Go faster*, or you'll be fined 50¢ during next period." [Second questionnaire]

Work Period 3—6 min. [base period for measuring change in production following fine]

Rest Period 3—5 min.; No Evaluation *S*s received no communication; No Fine *S*s received evaluation form, "Not fast enough. Won't be fined, but speed up." Fine *S*s received evaluation form, "Not fast enough. 50¢ fine. Won't be fined again, but speed up." [Third questionnaire]

Work Period 4—6 min. [at beginning of period, supervisor collected fine from pay containers of Fine *Ss*]

Rest Period 4—5 min. [at the end of this period the supervisor told all *Ss*, "I'm going out to find *E*, so he can give you his questionnaire. I suppose you might as well keep working for a while."] [Fourth questionnaire]

Work Period 5—4 min. [production in supervisor's absence; *E* stops work and distributes questionnaires when he returns]

Success of the Experimental Manipulation

Three assumptions about the experimental conditions must be stated and tested: (*a*) *Ss* perceived the threatened fine as a real punishment; (*b*) *Ss* perceived that the supervisor's coercive power existed, i.e., that he was capable of carrying out the fine; and (*c*) *Ss* in the 50¢ legitimate fine condition perceived the supervisor's evaluation, threat, and/or amount of fine as more legitimate than did the *Ss* in the 25¢ condition who were fined 50¢.

The data support the validity of the first assumption. On the post experimental questionnaire, *Ss* who were fined indicated that they would less "like to come back and work on this same task . . . under the same . . . conditions" ($p < .001$, by r_{bis} one-tailed) than *Ss* who were evaluated but not fined. The responses to two other questions were in the same direction but not significant:

Fine *Ss* "like[d] working on the task" less, and expected "people generally . . . to avoid [such] tasks" more, than did the nonfined *Ss*.

The second assumption is also supported by the questionnaire responses. Only two *Ss* reported after the task that the supervisor had no actual power to fine them.

The third assumption received rather weak support from the data. *Ss* in the 50¢ legitimate fine condition attributed more legitimacy to the supervisor on the following questions than did *Ss* in the 25¢ legitimacy condition:

Did you feel your supervisor's evaluations of your work were justified? ($p < .04$, by r_{bis} one-tailed); In terms of how you were doing on your job, do you think that the supervisor was justified in fining you at all? ($p < .06$, by r_{bis} one-tailed); Did you think the amount of the fine was fair? ($p < .04$, by r_{bis} one-tailed); *How much* would you have thought would have been a fair amount for the fine? ($p < .07$, by r_{bis} one-tailed).

Measures of Conformity

Measures were constructed to reflect *Ss*' conformity to the supervisor's influence attempts at three levels of observability.

1. *Most observable conformity.* Since *Ss* were told that the supervisor would count the total number of cards they sorted during each period, their "total production" was considered their most overt behavior. For a given work period, for example the fourth, the index of conformity to the speed-up induction was

$$\frac{S\text{'s Total Production in Period 4}}{S\text{'s Total Production in Period 3}}$$

henceforth abbreviated *T4/T3*.

2. *Less observable conformity.* Since *Ss* were told that the supervisor would not check the amount of *correct* production between periods, and because such checking appeared to be difficult, "correct production" was chosen as a measure of less overt conformity. For a given work period, for example the fourth, this index was

$$\frac{S\text{'s Correct Production in Period 4}}{S\text{'s Correct Production in Period 3}}$$

abbreviated *C4/C3*.

3. *Least observable conformity. Ss*' private evaluations of their performance were indicated on forms they filled out during each rest period "purely for the benefit of the experimenter" which it was impossible for the supervisor to see during the experiment. Thus, covert conformity was indicated by the degree to which *Ss* changed their self-evaluations in the direction of the supervisor's evaluation of their work. From the questionnaire filled out following the supervisor's evaluation, all *Ss* were classified as Accepting, Neutral, or Rejecting.

Measures of Resisting Forces

The measure of resisting forces was obtained on the postexperimental questionnaire from *Ss*' responses regarding "factors that tended to decrease your production."[3] *Ss* were asked to check as many of a set of alternatives as applied. Included in the list were these alternatives which were taken as resistance, since they referred to the influence attempt of the supervisor as the source of the opposing force:

I thought the supervisor's evaluation was unfair [Evaluation Unfair]

I resented the threat of being fined [Resent Threat]

I resented being fined [Resent Fine]

Fined *Ss*' responses to these three items were significantly associated as indicated in Table 5.

Measure of Punishment Forces

This index was based on three questions answered as part of the postexperimental questionnaire:

A. After he threatened to fine you, how much did you think the amount of the fine would be?

B. If you failed to comply with his recommendations, what did you think were the chances that the supervisor would fine you? [Expressed in probability.]

[3] Additional less direct indices of resisting forces were available from other questions, but only one ("Did you ever feel like getting up and walking out . . . ?") was significantly associated with the more direct measures. This single measure is not reported; in general it was not significantly related to the variables predicted.

C. Just after the supervisor evaluated your work for the first time, what did you think were the chances that you would be able to follow his recommendations enough so that you would avoid being fined? [Expressed in probability.]

The strength of the punishment force was computed by using A as the "valence of punishment," B as the "probability of punishment if nonconformity," and (1 − C) as the "probability of punishment if conformity" terms in the formula on p. 93.

HYPOTHESES AND RESULTS[4]

The first three hypotheses deal with the effects of the punishment force and resisting forces on conformity.

Hypothesis 1. The greater the strength of punishment force induced by O's threat the greater will be P's *conformity following the threat.* This hypothesis applies only to conformity during the time period 3/2 when there is still a possibility of future punishment. As shown in the first row of Table 2, there was a significant relationship in the predicted direction between strength of punishment force and conformity following threat. Hypothesis 1 was confirmed for all levels of observability, despite the fact that the questions from which the index of punishment force was computed referred only to total production.

These findings cannot be attributed to confounding of the punishment force measure with the experimental manipulation of legitimacy or with reported resisting forces. The means of the punishment force measure for *S*s in the two legitimacy conditions, and for *S*s with and without each resisting force, were nearly equal.

Furthermore, it can be derived that threatened *S*s (who, we assume, have a greater punishment force) would conform more than *S*s given an influence attempt without threat. However, since coercion also produces resistance (Hypothesis 5b was supported), this prediction could be counteracted by the resistance. Following the first influence attempt there were greater increases in threatened than in nonthreatened *S*s' mean total production ($p < .05$, two-tailed) and mean correct production ($p < .10$, two-tailed); however, there was no significant difference in acceptance of evaluation.

[4] Directional predictions were derived from the theory for all hypotheses, and therefore one-tailed tests of significance were employed. Where several factors led to doubt about the direction of prediction in supplementary analyses, two-tailed tests were used.

TABLE 2

RELATION BETWEEN STRENGTH OF PUNISHMENT FORCE[a] AND CONFORMITY IN PERIODS FOLLOWING THREAT AND PUNISHMENT

Time Period	Increase in Total Production	Increase in Correct Production	Acceptance of Evaluation
3/2 (following threat)	$r = +.44$[b] $p < .005$	$r = +.38$[b] $p < .01$	$r = +.35$ $p < .004$
4/3 (following punishment)	$r = +.09$[c] p *NS*	$r = -.37$[c] p *NS*[d]	$r = +.02$ p *NS*

[a] Strength of punishment force was computed from questionnaire responses, as explained on p. 96. The calculations are based on $N = 43$. Only those fined *S*s who answered the relevant questions, and whose punishment force index was nonnegative, were included.

[b] The correlation between strength of punishment force, and "differential conformity," i.e., increase in correct production relative to increase in total production following threat was +.09, indicating no significant difference in relation to these two measures of conformity.

[c] The correlation between strength of punishment forces, and increase in correct production relative to increase in total production following punishment was −.35 ($p < .05$, two-tailed), indicating a significant effect on differential conformity.

[d] In the wrong direction, as far as evidence for the *rejection of Hypothesis 2* is concerned.

Hypothesis 2. The strength of punishment force induced by O's threat is not positively associated with P's *conformity after the threatened punishment has terminated.* After punishment, conformity is no longer instrumental unless there is some expectation of further punishment, so the original punishment force should no longer influence *P*'s conformity. The results in the second row of Table 2 support the prediction made by Hypothesis 2 of no positive association. Again this relation is not confounded by the effects of resisting forces.

Similarly, Hypothesis 2 would predict that *S*s undergoing threat and punishment do not conform more following the fine than *S*s receiving influence attempts only. This comparison of fined and nonfined *S*s shows no significant differences in conformity following the administration of the fine.

Hypothesis 3. The greater the strength of resisting forces following an influence attempt, the less will be the conformity. The results for nonfined and fined *S*s were analyzed separately, since the test of Hypothesis 1 showed that

TABLE 3

MEAN CONFORMITY FOR *S*s WITH RESISTING FORCE ($\bar{X}_r$) MINUS MEAN CONFORMITY FOR *S*s WITHOUT RESISTING FORCE ($\bar{X}$)

Resisting Force Measure	*T3/T2* (Following Threat)		*T4/T3* (Following Punishment)	
	Increase in Total Production[a]	Increase in Correct Production[b]	Increase in Total Production[a]	Increase in Correct Production[d]
S reports production decreased because:				
Evaluation unfair	$\bar{X}_r - \bar{X} = -.03$ $t = -.87$ p *NS* $[\bar{X}_r - \bar{X} = -.10]$ $[t = -1.52]$	$\bar{X}_r - \bar{X} = +.06$ $t = +1.18$ p *NS*[e] $[\bar{X}_r - \bar{X} = -.11]$ $[t = -1.06]$	$\bar{X}_r - \bar{X} = +.02$ $t = +.45$ p *NS*[e]	$\bar{X}_r - \bar{X} = -.11$ $t = -2.00$ $p < .05$ $[\bar{X}_r - \bar{X} = -.06]$ $[t = -1.07]$
Resented threat	$\bar{X}_r - \bar{X} = -.06$ $t = -1.91$ $p < .05$ $[\bar{X}_r - \bar{X} = -.10]$ $[t = -1.55]$	$\bar{X}_r - \bar{X} = -.08$ $t = -1.50$ $p < .10$ $[\bar{X}_r - \bar{X} = -.15]$ $[t = -1.55]$	...	...
Resented fine	...	...	$\bar{X}_r - \bar{X} = -.03$ $t = -.87$ p *NS*	$\bar{X}_r - \bar{X} = -.12$ $t = -2.34$ $p < .05$ $[\bar{X}_r - \bar{X} = -.15]$ $[t = -3.06]$

[a] Based on 43 scores corrected for +.44 correlation with punishment force. Figures in brackets are based on uncorrected scores for all 55 fined *S*s.

[b] Based on 43 scores corrected for +.38 correlation with punishment force. Figures in brackets are based on uncorrected scores for 55 fined *S*s.

[c] Based on uncorrected scores for 54 fined *S*s (one walked out after the fine).

[d] Based on 43 scores corrected for −.37 correlation with punishment force. Figures in brackets are based on uncorrected scores for 54 fined *S*s (see Footnote c).

[e] Nonpredicted direction.

conformity was related to strength of punishment forces. In the former group no significant relationships were observed between the one available resisting force measure ("evaluation unfair") and conformity. In the Threatened-Fined group the covariance of production with strength of punishment force was removed where correlations were significant by computing the difference between conformity score and conformity predicted from punishment force. The means of (corrected) conformity scores for *S*s with and without each resisting force are compared in Table 3. Among threatened and fined *S*s Hypothesis 3 received fair support (apparently more strongly for correct than for total production), although the picture is unclear because of differing results for corrected and uncorrected scores (both of which are reported in Table 3).[8] An index of differential conformity was used to measure the ratio of observable conformity to less observable conformity. This $\frac{C4/C3}{T4/T3}$ index, based on uncorrected scores, showed that the difference in conformity for *S*s on the basis of the Resent Fine resisting force was significantly greater ($p < .02$, two-tailed) for *C4/C3* (i.e., covert) than for *T4/T3*. The resisting force measures were unrelated to *S*s' acceptance of evaluation.

Since there is some evidence that the legitimacy conditions differed in strength of resisting forces (see Hypothesis 6b), Hypothesis 3 would predict that conformity would be greater in the Legitimate than in the Illegitimate Fine condition. However, there were no significant differences using either corrected or uncorrected production scores or the measure of acceptance of evaluation, probably because the differences in resistance were too small.

The last three hypotheses deal with attraction, coercion, and legitimacy as determinants of resisting forces.

[8] Although observed effects could not be attributed to differences in punishment forces, scores were corrected in the hope of decreasing extraneous variance.

Hypothesis 4. The strength of P's *resisting forces against an influence attempt will vary inversely with his attraction to O.* This hypothesis states that resisting forces are related to *P*'s "dislike" of *O*. *P*'s attraction to *O* was measured by *Ss*' responses to the question, "If you had a chance to work for your supervisor on some other part-time or summer job, do you think you would work for him?" (Responses to a more direct measure of attraction, namely "liking," were practically invariate.)

Hypothesis 4 received moderate support from the data. For *Ss* in the Fine conditions there were inverse associations between attraction and individual resisting force measures for "evaluation unfair," "resent threat," and "resent fine" of $-.30$ ($p < .02$), $-.11$ (p *NS*), and $-.23$ ($p < .04$), respectively as measured by biserial tau. The association between the total number of such resisting forces and the measure of attraction was $r = -.18$ ($p < .04$, $N = 55$). Within the No Fine condition the association between the one resisting force measure and attraction was approximately zero ($N = 21$).

It should be observed that the hypothesis does not test the causal dependence of resisting forces on the relationship with the supervisor.

Hypothesis 5a. The strength of attraction toward O *decreases with threat and punishment.* As predicted, it was found that *Ss* in the Legitimate Fine condition were less attracted to *O* (measured in the same way as in Hypothesis 4) than those in the No Fine condition ($r_{bis} = +.39$, $p < .001$, $N = 54$). This result confirms the previous finding by Raven and French (1958b).

Hypothesis 5b. From Hypotheses 5a and 4 we derive: *the strength of resisting forces increases with threat and punishment.* As indicated in Table 4, *Ss* in the No fine condition reported resisting forces less often than *Ss* in the Legitimate Fine condition. The difference for resisting forces due to evaluation was not significant. Although there was a highly significant difference for resistance due to the fine, the measure of resistance due to the fine was not entirely appropriate in the No Fine condition where the *Ss* received only an indirect threat and were never fined. A more appropriate comparison would have been possible only if the control group received a very small fine instead of no fine at all. Thus the findings are consistent with the predictions and the results from Hypotheses 4 and 5a, but the test of the hypothesis is not very satisfactory.

TABLE 4

PROPORTION OF *Ss* REPORTING RESISTING FORCES WITHIN EXPERIMENTAL CONDITIONS

Resisting Force Measure	Experimental Condition		
	No Fine ($N = 21$)	Legitimate Fine ($N = 33$)	Illegitimate Fine ($N = 22$)
S reports production decreased because:			
Evaluation unfair	.24	.27	.55
Resent threat	(.19)[a]	.39	.50
Resent fine	(.05)[a]	.52	.64

[a] *Ss* in the No Fine condition received a less direct threat than those in the other groups, and were never fined.

Hypothesis 6a. The strength of attraction toward O *increases with increasing legitimacy of an influence attempt.* Especially where coercion is used, the legitimacy of *O*'s acts should lessen the decrease in attraction which threat and punishment presumably cause. Contrary to the prediction, *Ss* in the Legitimate Fine condition were less attracted to *O* as compared with *Ss* in the Illegitimate Fine condition ($r_{bis} = +.08$, NS) A further test of the hypothesis by relating attraction to *questionnaire* measures of legitimacy also failed to confirm it (see Table 5). This finding fails to support the two previous experiments by Raven and French (1958a; 1958b), though we are here dealing with legitimacy of punishment rather than legitimacy of role occupancy.

Hypothesis 6b. From Hypotheses 6a and 4 we derive: *the strength of resisting forces decreases with increasing legitimacy.* However, since 6a was not confirmed, we can no longer assume that the effect of legitimacy of punishment on resisting forces is mediated by attraction. The last two columns in Table 4 show that the illegitimate treatment produced more resisting forces by all three measures, but only the "evaluation unfair" measure is significantly different in the two experimental conditions ($p < .05$, χ^2 with Yates' correction). Biserial tau between the total number of resisting forces and the two experimental conditions was $+.20$ ($p < .07$).

A further examination of the evidence for

TABLE 5
ASSOCIATIONS BETWEEN MEASURES OF LEGITIMACY, ATTRACTION, AND RESISTING FORCES
(For fined *Ss*)

Variables		Illegitimacy					Resisting Forces (*S* Reports Production Decreased Because)		
		1	2	3	4	5	6	7	8
Variables	Measure	Evaluations of Work Unjustified	Unjustified in Fining at All	Amount of Fine Unfair	Fair Amount of Fine (Small)	Attraction to supervisor	Evaluation Unfair	Resent Threat	Resent Fine
	N =	55	55	46*	39*	55	55	55	55
2	r =	+.74	—	—	—	—	—	—	—
	p <	.001							
3	r =	+.65	+.67	—	—	—	—	—	—
	p <	.001	.001						
4	r =	+.51	+.59	+.59	—	—	—	—	—
	p <	.001	.001	.001					
5	r =	−.03	+.01	+.01	+.01	—	—	—	—
	p <	*NS*	*NS*	*NS*	*NS*				
6	r_{bis} =	+.45	+.36	+.54	+.43	−.30	—	—	—
	p <	.001	.003	.001	.003	.02			
7	r_{bis} =	+.08	+.10	+.39	+.29	−.11	+.37	—	—
	p <	*NS*	*NS*	.003	.03	*NS*	<.008		
8	r_{bis} =	+.24	+.26	+.31	+.08	−.23	+.32	+.40	—
	p <	.03	.03	.02	*NS*	.04	.03	.004	

* These measures were not obtained from some *Ss* early in the experiment; hence the *N*s for Variables 3 and 4 are less than 55.

Hypothesis 6b, utilizing measures of association between the resisting force questions and the *questionnaire* measures of illegitimacy used in the validation of the legitimacy manipulation (see Method section), indicated significant positive associations. As shown in Table 5, 9 of the 12 biserial tau's were significant beyond the .05 level. These highly significant associations should be regarded with caution, however, because the illegitimacy and resisting force questions were similar in content, although the latter questions appeared in a separate portion of the questionnaire in which *Ss* were asked to attribute decreases in production to various causes; in addition the distinction between questionnaire measures of legitimacy and resisting forces is supported by their differential association with attraction (see Table 5).

On the basis of the evidence, especially the former tests which utilized the experimental conditions of legitimacy and illegitimacy, Hypothesis 6b receives some support. Comparing the results of Hypotheses 6b and 4, it can be seen that the measure of attraction and the experimental manipulation of legitimacy are similarly related to the resisting force measures, although unrelated to each other.

SUMMARY

In a laboratory experiment 90 *Ss* worked on a card sorting task under a supervisor who created various conditions of coercive power. Thirty-three (Legitimate Fine) *Ss* were threatened for working too slowly and later fined the amount they had been told was legitimate; 22 (Illegitimate Fine) *Ss* were threatened and fined the same amount which, in this condition, was twice the specified legitimate amount of fine; and 21 (No Fine) *Ss* were neither threatened nor fined although they were told they were working too slowly.

It was hypothesized that the supervisor's inductions would set up punishment forces and opposing resisting forces both of which would determine *S*'s conformity. Assuming that conformity to coercive power is instrumental to avoiding punishment, the strength of the punishment force should increase with in-

creasing negative valence of the fine and with increasing probability of avoiding it by conforming. The valence and probability of receiving a fine were measured on a questionnaire. The resisting forces should be determined by attraction to the supervisor (i.e., disliking) which in turn is determined by coercion and the legitimacy of coercion.

The findings supported the following hypotheses:

1. The greater the strength of punishment force induced by *O*'s threat, the greater will be *P*'s conformity following the threat.

2. The strength of punishment force induced by *O*'s threat is not positively associated with *P*'s conformity after the threatened punishment has terminated.

3. The greater the strength of resisting forces following an influence attempt, the less will be the conformity.

4. The strength of *P*'s resisting forces against an influence attempt will vary inversely with his attraction to *O*.

5a. The strength of attraction toward *O* decreases with threat and punishment.

6b. The strength of resisting forces decreases with increasing legitimacy.

The following hypothesis was not directly tested, but the data presented were consistent with the prediction as were the findings for Hypotheses 4 and 5a.

5b. The strength of resisting forces increases with threat and punishment.

The findings did *not* support the following hypothesis:

6a. The strength of attraction toward *O* increases with increasing legitimacy of a coercive influence attempt.

For all of these hypotheses except Number 4 the tests included comparisons among experimentally manipulated conditions and hence permitted an unequivocal distinction between the independent and the dependent variables. For Hypothesis 3 only the correlational evidence was significant.

It is concluded that punishment forces and resistance forces are independent determinants of conformity to coercive power. Although it appears that legitimacy affects resisting forces, the present experiment indicates that such effects are not mediated by low attraction to the supervisor.

REFERENCES

FRANK, J. D. Experimental studies of personal pressure and resistance. *J. gen. Psychol.*, 1944, **30**, 23–64.

FRENCH, J. R. P., JR., & RAVEN, B. The bases of power. In D. Cartwright (Ed.), *Studies in social power*. Ann Arbor, Mich.: Inst. Soc. Res., 1959. Pp. 150–167.

RAVEN, B., & FRENCH, J. R. P., JR. Group support, legitimate power, and social influence. *J. Pers.*, 1958, **26**, 400–409. (a)

RAVEN, B., & FRENCH, J. R. P., JR. Legitimate power, coercive power, and observability in social influence. *Sociometry*, 1958, **21**, 83–97. (b)

(Received July 20, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 102-109

RESISTANCE AND CONFORMITY UNDER REWARD AND PUNISHMENT[1]

SHEILA G. ZIPF
University of California, Berkeley

THE PROCESS of social influence has received considerable study in recent years. Findings indicate that conformity varies with such factors as the characteristics of the influencer, the methods of influence used, the nature of the behavior affected, characteristics of the person receiving the influence and, the relationship between the influencer and the recipient (Asch, 1956; Carlson, 1956; Festinger, 1957; Frank, 1944; Hovland, Janis, & Kelley, 1953; Lippitt, Polansky, Redl, & Rosen, 1953). Two aspects of this area seem to have been neglected, however. One of these is the influence process as it affects *actions*. A majority of the studies have been concentrated in influence on opinions or judgments. The other is systematic consideration of the psychological processes that intervene between external conditions and the conformity behavior to produce differential conformity.

Let us view influence on action as a process in which the motivational systems of *Individual P* are changed by the act of another *Person O*. The process of change in motivation intervenes between the influence act of *O* and the conformity behavior of *P*. In terms of individual motivation, one can readily conceive that a single stimulus event may produce conflicting motivations. Thus, the act by which *O* attempts to influence *P* may not only motivate *P* to conform to *O*'s wish, but may also motivate *P* to resist doing what *O* wants. French and Raven (1959) have distinguished several types of influence and proposed that these types should be studied for their differential effects in motivating the person to resist conforming as well as their effects in motivating the person to conform. The terms Conformity Force and Resistance Force will be used for these two conflicting tendencies caused by an influence attempt. French, Morrison, and Levinger (1960) following Lewin (1938) have suggested a formula for the Conformity Force which combines the valence of the goal and the probability of reaching the goal if one does try to conform.

The present study was designed to investigate some of the factors involved in producing the conformity and resistance forces and the resulting changes in action. In order that two different acts would not be involved, a situation was selected in which the conformity force would lead to an increase in a given action and the resistance force to a decrease in the same action. Because the effect of the resistance force was of central interest, it was necessary to produce conditions for which the conformity force could be equated while the resistance force was allowed to vary.

Rewards and punishments, being basic types of influence, were of primary interest. Furthermore, because monetary rewards and punishments permit a reasonable equating of the formula for the strength of the conformity force and were expected to produce different strengths of resistance force, these two types of influence were selected for comparison.

Some of the fairly stable motivational structures of the person should also affect the extent to which an influence attempt on that person produces a resistance force. Since the personality characteristic most likely to be relevant seemed to be the need for independence from control figures, this was included as a variable in the study.

With valence defined as the attractiveness of the goal for the person being influenced and perceived probability as the probability perceived by the person being influenced, the following postulates were framed.[2]

1. The strength of the force to do Action X

[1] This paper is based on a thesis submitted in partial fulfillment of the requirements for the PhD degree at the University of Michigan. The research was done under contract with the Office of Naval Research at the Research Center for Group Dynamics. Part of this study was reported at the APA meeting in Washington in August 1958. The author wishes to express her appreciation to John R. P. French, Jr. for his valuable assistance in every stage of the study, to John W. Atkinson and Charles Hubbel for many theoretical suggestions, and to Cathryn A. Levison for her constructive review of the manuscript.

[2] A complete presentation of the theory is given in the original report (Zipf, 1958).

is an increasing function of the valence of the Goal G, where Action X leads to Goal G.

2. The strength of a force to do Action X is an increasing function of the perceived probability that Action X leads to Goal G.
3. The strength of a force to do Action X is an increasing function of the perceived probability that one is able to perform Action X in order to reach Goal G.

These general postulates were used to make derivations for both the conformity and the resistance force.

Briefly, the derivations concerning type of influence and personality as they affect the resistance force were as follows:

The threat and administration of a punishment should tend to make the entire situation negatively valent. There should be forces to leave the field. These forces should operate as resistance forces, opposing the forces to remain in the field and perform Action X. This should not occur in response to the offer and administration of rewards. For those with strong needs to be independent, conforming to the demands of a control figure should itself be somewhat negatively valent. Therefore, there should be a force to not conform. Those who do not have such strong needs for independence should have a less negative valence, or even a positive valence on conforming. Hence, two factors were expected to increase the resistance force and in turn decrease conformity (*a*) punishment and (*b*) the need for independence.

METHOD

Subjects

One hundred and two female *S*s were recruited from undergraduate classes at the University of Michigan. The measure of need for independence was obtained at this time, by a different person who said he was not related to the study but was just pretesting the instrument. He asked everyone to fill out the questionnaire regardless of whether they were signing up for the experiment or not. *S*s were assigned to the laboratory in groups of two to eight so as to keep the distribution of need for independence scores about the same in each condition. Otherwise, assignment was random.

Task

The task the *S*s were given was to sort IBM cards according to the total number of holes they contained. A supervisor, who was said to be connected with a different project for which the cards had to be sorted, was introduced. She explained that she was interested in getting the job done and that the only effect of the experiment was that the *E* was allowed to give out and collect his questionnaires during rest periods and because he wanted conditions of low communication, she had agreed to make any necessary comments to them in the form of written notes. The work was done in booths. There were five work periods of 4 min. each. A more detailed description of the work situation is given in French, Jr., & Morrison, & Levinger (1960).

Manipulation of Sanction

The supervisor explained that she could not make any check on their accuracy during the session, but by measuring the height of the deck of cards they had sorted could tell the speed at which they performed. She indicated that after each work period she would collect their production, evaluate it for speed and give them an evaluation sheet with her recommendations on it. The evaluations were to be made by comparing their work with that of regular employees on the task. Half of the *S*s were told that they might be fined if their work was not up to standards, but that they would be warned of this possibility before the work period in which it would occur. The other half of the *S*s were told the same thing concerning rewards. All *S*s were told that they would start out with a $1.00 payment which was placed in advance in each of the booths. The sequence of events after *S*s had gone to their booths was as follows:

Work Period 1
Rest Period—Cards collected; *S*s told that was a practice period
Work Period 2
Rest Period—Cards collected
Evaluation: Much too slowly
Reward induction: Go faster next time and there will be a 50¢ reward.
Fine induction: Go faster next time or there will be a 50¢ fine.
Self evaluations made
Work Period 3
Rest Period—Cards collected; conformity and resistance forces measured
Evaluation—Too slowly
Reward induction: 50¢ reward for improvement, 50¢ reward next time if you go faster. [reward given]
Fine induction: Improvement, but not fast enough, 50¢ fine. Go faster next time or there will be a 50¢ fine. [fine taken]
Self evaluations made
Work Period 4
Rest Period—Cards collected; conformity and resistance forces measured
Evaluation: A little too slowly
Reward induction: 50¢ reward for improvement. No reward next time, but go faster. [reward given]
Fine induction: Improvement, but not fast enough, 50¢ fine. No fine next time, but go faster. [fine taken]
Self evaluations made
Work Period 5
Rest Period—Cards collected; resistance force and

expectation of sanction measured
Evaluation: A little too slowly
Self evaluations made
Final questionnaire administered

Concepts and Measures

The conformity force is defined and measured as valence of the goal, times the probability of reaching the goal by attempting to conform.[3]

Explicitly:

The desire to receive the reward or avoid the fine $\times$

The perceived probability that one will be fined or fail to obtain the reward if he does not conform $-$ The perceived probability that one cannot conform enough to avoid the fine or obtain the reward

These components of the conformity force were measured by asking the *S*s how much they wanted to avoid the fine or obtain the reward, and what they thought the probabilities were, on fixed choice scales.

The resistance force is conceived as having the same general form as the conformity force. However, no measures were found suitable to that form and the resistance force was measured in a checklist of "factors leading to lowered production." Items that referred to the behavior of the supervisor, such as "I didn't like the way the supervisor acted," were used. In a list of 22 items, there were 7 such items and the number checked was used as the index of the resistance force.

Conformity was measured by the change in speed following an induction to go faster. The index was: total number of cards sorted in the work period following the induction, divided by the total number of cards sorted in the work period prior to the induction. This index is abbreviated $T\ n/n-1$ where n is the number of the work period for which the induction to go faster was given.

A measure of judgmental conformity was also included. This was on the Evaluation Sheet, on which the supervisor evaluated the *S*'s work on a seven point scale from "much too slowly," to "much too fast." All *S*s received the same evaluation for a given work period. On the same page, *S*s were asked to give their own evaluation of their work on the same scale. The difference between the supervisor's check and the *S*s check is the index of disagreement.

The need for independence was measured by 6 items in a questionnaire. Originally 12 items were selected from a test used by Tannenbaum and Allport (1956) and given to 60 college women. The correlation matrix of the items was obtained and 6 items which clustered and best met the definition of the need to be independent of control figures were used. Since the hypotheses concerning the effects of this need are made for those with a strong need, *S*s referred to as High NI are in the top one third of the experimental *S*s on this measure. Those called Low NI are in the lower two thirds. In all other measures, a break of the distribution into high and low is made as close to the median as possible.

Effectiveness of Manipulations

The main assumption concerning the manipulation of sanctions is that the situation is generally less attractive for the Fine *S*s than for the Reward *S*s. This was tested in several ways. On the question, "How do you feel about the task?" Fine *S*s like the task less than Reward *S*s in each work period for which it was asked (Work Periods 2, 3, 4, 5, $P \leq .05, .06, .05, .08$ respectively, χ^2 test). When asked at the end of the experiment if they ever felt like getting up and walking out, Fine *S*s said they did more often than Reward *S*s ($P \leq .01$, χ^2 test). The assumption appears to be valid.

Another assumption is that the rewards and fines of the same amount may be equated. We found that Fine *S*s do not differ in the degree to which they said they wanted to avoid the fine from the degree to which Reward *S*s said they wanted to obtain the reward, at either time that the question was asked ($P > .30$ one tail t test).

HYPOTHESES AND RESULTS

Hypothesis 1.

The conformity force is positively related to increase in speed following an induction to go faster. This hypothesis is clearly confirmed in data from the fourth work period.[4] The correlations are shown in Table 1. Also in that table, the correlations of each of the components of the conformity force with conformity are given. All of the components are related in the appropriate direction; valence, significantly so. The conformity force is better

[3] This specific formula was suggested by French et al. 1960. It is similar both in variables used and in the multiplicative form to the formulations of a number of theorists for the prediction of choice and intensity of behavior (Atkinson, 1957; Lewin, Dembo, Festinger, & Sears, 1944; Rotter, 1954; Siegel, 1957; Tolman, 1955). The probabilities as given in the formula differ slightly from the formal statement of the components of the force in the postulates. This is primarily due to the necessity of stating questions so they could be readily understood by the *S*s.

[4] The hypotheses were generally not confirmed by data from the third work period. A previous study (French, 1960) had obtained some of the predicted results under similar conditions in the third work period. However, 6-min. work periods were used at that time. It appears that a certain amount of practice on the task is necessary before these motivational variables are effective. Only the fourth and fifth work period results will be given in this report. There are, also ten *S*s omitted from the fourth work period data. Their conformity force measures indicated that they did not understand the questions. They have been omitted from the data on the fourth work period for clarity of presentation. In no case does the exclusion of these 10 change whether the result of an hypothesis is significant or not although the magnitude of the effects are generally greater than when they are included.

TABLE 1

CORRELATIONS OF THE CONFORMITY FORCE VARIABLES AND THE CONFORMITY FORCE FORMULA WITH INCREASE IN SPEED IN THE FOURTH WORK PERIOD

(T 4/3)

Variable	r	N	p
A. Valence of the goal	.31	92	.001
B. Probability that one will be fined or fail to receive the reward for nonconformity	.16	92	.06
C. Probability that one cannot conform enough to avoid the fine or obtain the reward	−.17	92	.06
Conformity Force Formula A[B − C]	.44	92	.001
Conformity Force Formula *Ss* have Low Resistance Force	.36	63	.001
Conformity Force Formula *Ss* have High Resistance Force	.41	29	.01

Note.—All p values given are for one-tail tests.

TABLE 2

RELATIONSHIP OF THE RESISTANCE FORCE TO CHANGE IN SPEED OF SORTING IN THE FOURTH WORK PERIOD

(T 4/3)

		Resistance Force		
		Low	High	
All *Ss* (N = 92)				
	Low	27	20	
T 4/3				
	High	36	9	$\chi^2 = 6.42$
				$p = .02$

		Resistance Force		
		Low	High	
Low Conformity Force (N = 46)				
	Low	16	13	
T 4/3				
	High	15	4	$\chi^2 = 1.66$
				$p = .10$

		Resistance Force		
		Low	High	
High Conformity Force (N = 44)				
	Low	11	7	
T 4/3				
	High	21	5	$\chi^2 = 1.20$
				$p = .15$

Note.—All p values given are for one-tail tests.

related than any of its three components, and is significantly as predicted for the total group. Partial correlations of each of the components of the conformity force on T 4/3 have been obtained for the Reward *Ss* only. They indicate that each variable is significantly ($p < .01$) related to change in speed at that time, in support of the postulates given above.

Since the conformity force and the resistance force are conceptually similar and are predicted to have opposite relationships to conformity, one can be sure that a result obtained with one is not due to the other only if it can be shown that the two are independent. In the fourth work period, the two forces are not significantly related ($p > .50$, χ^2 test). To remove any possibility of confounding, however, when presenting the results of one of the forces, the other is controlled by holding it at two levels; high and low. This is not complete control. However, since the measure of resistance force is extremely skewed, over 50% of the *Ss* having a score of zero, other methods cannot be used. As shown at the bottom of Table 1, the first hypothesis is confirmed when this control is used.

Hypothesis 2.

The resistance force is negatively related to increase in speed after an induction to go faster. These data for the fourth work period are presented in Table 2. The relationship is significant for the total N, but not so when the conformity force is held at high and low. This finding may indicate that the relationship in the total N is caused by the conformity force. However, since the conformity and resistance forces are not related, it seems more likely that the reduction in N caused by the control is responsible for the loss of significance.

The assumption involved in measuring the conformity and resistance forces that have opposite relationships to conformity is that the resultant of the two should predict extent of conformity better than either alone. Because the two forces were not measured in the same units, the resistance force could not simply be subtracted from the conformity forces as would be preferred. Furthermore, the resistance force distribution is not suitable for partial regression analysis. But a simple combination of the two forces is possible. Dichotomizing the distributions, *Ss* were grouped according to whether the conformity and resistance forces were both high, both low, or opposite to each other. The relationship of such a combination to increase in speed

TABLE 3

RELATIONSHIP OF THE CONFORMITY AND RESISTANCE FORCES TO CHANGE IN SPEED OF SORTING IN THE FOURTH WORK PERIOD

(T 4/3)

	Conformity Force / Resistance Force	High Low	High High	Low Low	Low High	
T 4/3	Low	11	7	16	13	
	High	21	5	15	4	$\chi^2 = 8.22$ $p < .05$ (two-tail test)

TABLE 4

COMPARISON OF REWARD AND FINE CONDITIONS ON MEAN RESISTANCE FORCE IN THE FOURTH AND FIFTH WORK PERIODS

Fourth Work Period	Fine		Reward		
	N	$\bar{X}$	N	$\bar{X}$	t
All *Ss*	40	1.15	52	.25	3.82[a]
Low Conformity Force	22	.91	27	.35	1.83[a]
High Conformity Force	18	1.44	25	.15	3.59[a]
Fifth Work Period					
All *Ss*	47[b]	.77	55[b]	.09	4.50[a]
No Sanction Expected	29	.83	36	.01	15.04[a]
Sanction Expected	9	.56	13	.23	2.99[a]

[a] $p \leq .05$, one-tail test when correction for number of degrees of freedom is made because of a significant difference in variances (Dixon & Massey, 1951).

[b] The total N is greater than the sum of N in the two subgroups because expectation of sanction was not ascertained for some *Ss*.

is given in Table 3. When the two forces are opposite to each other, the change in speed is as expected. When both forces are high or low, the increase in speed is as likely to be high as low. This finding lends support to the idea of a resultant force combining the conformity and resistance forces to predict extent of conformity and also makes acceptance of Hypothesis 2 more credible.

Hypothesis 3.

Resistance forces are greater in Fine than in Reward. In the fourth work period, this hypothesis is clearly confirmed, as shown in the top half of Table 4.

Hypothesis 4.

Conformity is greater in Reward than in Fine. This hypothesis was not confirmed. There are no significant differences between the Reward and Fine conditions on increase in speed of sorting at this time, neither when the conformity force is held at high and low, nor with the total N. The hypothesis follows logically from the theory only if Hypotheses 1, 2, and 3 are substantiated. That Hypothesis 2 received questionable confirmation may therefore explain this result. Another factor is that only 47% of the Fine *Ss* had a score of greater than zero on the resistance force measure and the mean for them alone was only 2.56 out of a possible 7. It is likely that the resistance forces which were produced by the fines tended to have the predicted effects as indicated by Tables 2, 3, and 4, but that it was not produced in sufficient strength in enough of the Fine *Ss* to result in significant differences in conformity between the Fine and Reward conditions.

A secondary test of Hypothesis 4 may be made concerning judgmental conformity, using the measure of disagreement with the supervisor's evaluation. It was found that Fine *Ss* disagree with the supervisor more than do the Reward *Ss* ($p = .05$, one-tail t test). The *Ss* knew that their self-evaluations would not be seen by the supervisor and could not be sanctioned. On the other hand, speed on the task was specifically being sanctioned. Two possible conclusions are suggested: (*a*) Hypothesis 4 is correct only with respect to judgmental conformity and not for action conformity; (*b*) the resistance force generated by this situation does lower conformity of Fine *Ss* below that of Reward *Ss*, but only when the nonconformity does not put the resistor in danger of being punished.

The fifth work period was carried out to observe the effects of the resistance and conformity forces after sanctions had been removed. It was expected that Fine *Ss* would still have resistance motivation and tend to conform less than Reward *Ss* even though there would be no more fines or rewards.

The conformity force was not measured for this work period; instead, a measure was taken of whether or not *Ss* expected that there would be sanctions at this time. This measure is not related to conformity in this work period, directly, but it does have an indirect effect that will be discussed presently. The same measure of resistance that was used in the previous work period was given again for the

TABLE 5
COMPARISON OF REWARD AND FINE CONDITIONS ON MEAN INCREASE IN SPEED OF SORTING IN THE FIFTH WORK PERIOD
(T 5/4)

	Fine		Reward		
	N	$\bar{X}$	N	$\bar{X}$	t
All *Ss*	47[a]	1.00	55[a]	1.00	0
No Sanction Expected	29	.99	36	1.04	2.92[b]
Sanction Expected	9	1.01	13	.99	−.37[c]

[a] The total N is greater than the sum of N in the two subgroups because expectation of sanction was not ascertained for some *Ss*.
[b] $p \leq .01$, one-tail test.
[c] A negative t indicates a difference in the nonpredicted direction.

fifth work period. The measure is not related to conformity at this time. Its validity depends on whether the checklist contains the appropriate sources of resistance. The measure seems to operate as expected with fourth work-period data, but it is likely that the factors producing resistance motivation in the fourth work period were different from those in the fifth work period, since the situation had changed considerably. The items used, apparently, were not very relevant to conformity on the task in the fifth work period.

The resistance force index does seem to indicate a negative attitude at this point, however, because it again turns out to be higher for Fine *Ss* than for Reward *Ss*. This is shown on the lower half of Table 4 for the total group, with expectation of sanction held constant.

The experimental groups were compared on conformity for this work period. Again, those who expected sanctions were separated from those who did not. This comparison is given in Table 5. When sanctions were not expected, Fine *Ss* conform less than the Reward *Ss*. When sanctions are expected, the Reward and Fine *Ss* do not differ in conformity.

The results of the fifth work period seem to indicate that some amount of resistance force persisted in Fine *Ss* even when there were no more fines threatened. The conformity finding lends support to the second conclusion offered for the results of Hypothesis 4 above. In this work period, it was possible to compare two sets of Reward and Fine groups, one where sanctions were still expected and the other where they were not. For those who do not expect sanctions, resistance evidenced by slowing down on the task does not endanger the resistor with punishment or loss of reward. For those who do expect sanctions, slowing down does endanger the resistor in this way. In the fifth work period when the danger is not present the Fine *Ss* conform less than the Reward *Ss*, however, when the danger is perceived to be present, the conformity of the Reward and Fine *Ss* does not differ.

Figure 1 shows that a similar conclusion might have been made by considering simply the number of cards sorted by each of the experimental groups in each work period. Although the Reward *Ss* tend to perform at a higher level throughout, there is a significant difference only after sanctions had ceased, in the fifth work period ($P < .04$, one-tail t test).

A number of questions were asked on the questionnaire given at the end of the experiment. The responses are of interest because they indicate the effects of the experimental manipulation on factors other than conformity, particularly those motivations which were not explicitly included in the measures of the conformity and resistance forces. Each effect mentioned was tested on a two-fold contingency table ($p < .05$, χ^2 test). Some of the findings are that the Fine *Ss* liked the task, the situation and the supervisor less than did

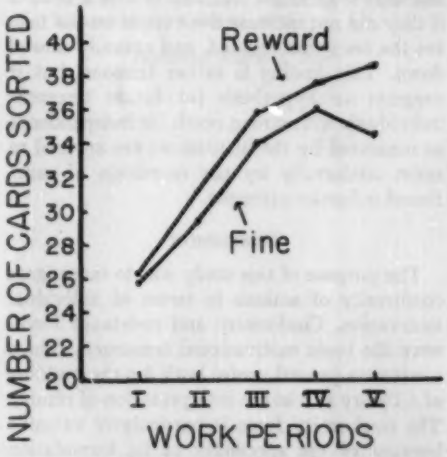

FIG. 1. Speed of sorting of Reward and Fine *Ss* in each work period.

the Reward *S*s. They felt the supervisor was more annoying, less competent, and her evaluations less justified. They thought it was less important to the supervisor that they do well on the task, and they felt less accomplishment from going faster. The rewards and punishments thus appear to have affected the satisfactions of a number of motives, such as achievement, aggression, and affiliation, as well as the need for independence, on which more explicit data are available.

It was predicted that, regardless of the experimental condition, those with High NI would have greater resistance forces and less conformity than those with Low NI. This measure does not differentiate individuals on resistance or conformity in the fourth work period. However, in the fifth work period, the difference in the resistance force approaches significance ($p = .07$, one-tail t test, $N = 102$). Comments from *S*s on their questionnaires and in interviews later indicate that the index of resistance force would have been more sensitive to the reactions of *S*s with High NI had it included items referring to the incessant character of the supervisor's demands.

The same comparison was made on conformity of action (T 5/4). Those who are High NI tend to conform less than those who are Low NI but the difference is significant only for those who expect sanctions ($p < .05$, one-tail t test, $N = 22$). *S*s with High NI who felt that they might be fined or lose a reward if they did not increase their speed on the task are the ones who resisted, and actually slowed down. This finding is rather tenuous, but it suggests an hypothesis for future research; individuals with strong needs for independence as measured by the instrument are aroused to resist conformity by the repetition of sanctioned influence attempts.

DISCUSSION

The purpose of this study was to investigate conformity of actions in terms of individual motivation. Conformity and resistance forces were the basic motivational constructs. These constructs proved useful both for the building of a theory and in the interpretation of results. The conformity force is particularly valuable because of the specificity of its formulation and its predictive ability. It would seem to offer an excellent framework for more basic definitions of different kinds of influence. Types could be defined in terms of the components of the conformity force which are changed and the extent of the change.

The resistance force, though less adequately measured, does have explanatory and predictive value. The weak test made of the effectiveness, as a predictor of conformity, of a resultant force combining the conformity and resistance forces received reasonable support.

More work is needed to improve the measure of the resistance force so that it is composed of the postulated components of a motivational force. This may be done by considering all of the relevant goals in the situation and summing all the forces that are opposite to the conformity force (Carlson, 1956; Thomas & Zander, 1959). An alternative is the way that the conformity force was handled in the present study. In this case a single goal, reward or avoidance of fine, served as a symbolic goal for many nonmonetary goals simplifying the measure. Such improvements in the measure of the resistance force would permit further investigation into the conditions under which varying strengths of the resistance force do or do not produce actions and judgements that are contrary to the demands and judgements of the influencing person.

SUMMARY

An experiment was designed to compare influence by means of rewards and punishments for differential effects on motivation to conform (conformity force) and motivation to resist (resistance force) and the resulting change in actions. Subjects were premeasured on the need to be independent of control figures. The following hypotheses were confirmed:

1. The conformity force taken as a combination of valence of the goal and the probability that trying to conform will lead to the goal, is positively related to conformity of action as measured by increase in speed following a sanctioned induction to go faster.

2. The resistance force is negatively related to conformity of action as measured by increase in speed after a sanctioned induction to go faster.

3. Influence by means of punishment pro-

duces a greater resistance force than influence by means of reward.

The following hypotheses for further research were suggested:

1. When the amount of punishment and reward are equal, persons influenced by punishment conform less than those influenced by reward if the nonconformity behavior is not likely to lead to punishment. This relationship does not hold for nonconformity behavior that is likely to lead to punishment.

2. The motivation to resist produced in those influenced by means of punishment persists for some period of time after punishments have ceased.

3. Persons with strong needs to be independent of control figures are aroused to resist conforming by repeated influence attempts involving sanctions.

REFERENCES

ASCH, S. E. Studies of independence and conformity: I. A minority of one against a unanimous majority, *Psychol. Monogr.*, 1956, **70**(9, Whole No. 416).

ATKINSON, J. Motivational determinants of risk-taking behavior. *Psychol. Rev.*, 1957, **64**, 359–372.

CARLSON, E. S. Attitude change through modification of attitude structure. *J. abnorm. soc. Psychol.*, 1956, **52**, 256–261.

COCH, L., & FRENCH, J. R. P., JR. Overcoming resistance to change. *Hum. Relat.*, 1948, **1**, 512–532.

DIXON, W., & MASSEY, F. *Introduction to statistical analysis*. New York, McGraw-Hill, 1951.

FESTINGER, L. *A theory of cognitive dissonance*. Evanston, Ill.: Row, Peterson, 1957.

FRANK, J. D. Experimental studies in personal pressure and resistance: I. Experimental production of resistance. *J. gen. Psychol.*, 1944, **30**, 23–41.

FRENCH, J. R. P., JR., & Raven, B. The bases of social power. In D. Cartwright, (Ed.). *Studies in social power*. Ann Arbor, Mich.: Inst. Soc. Res., 1959. Pp. 150–167.

FRENCH, J. R. P., JR., MORRISON, H. W., & LEVINGER, G. Coercive power and forces affecting conformity. *J. abnorm. soc. Psychol.*, 1960, **61**, 93–101.

HOVLAND, C. I., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion*. New Haven: Yale Univer. Press, 1953.

LEWIN, K. The conceptual representation and the measurement of psychological forces. *Contr. psychol. Theor.*, 1938, **1**, No. 4.

LEWIN, K., DEMBO, TAMARA, FESTINGER, L., & SEARS, PAULINE S. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944.

LIPPITT, R., POLANSKY, N., REDL, F., & ROSEN, S. The dynamics of power. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory*. Evanston, Ill.: Row, Peterson, 1953.

MACCORQUODALE, K., & MEEHL, P. Preliminary suggestions as to a formalization of expectancy theory. *Psychol. Rev.* 1953, **60**, 55–63.

ROTTER, J. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.

SIEGEL, S. Level of aspiration and decision making, *Psychol. Rev.*, 1957, **64**, 253–262.

TANNENBAUM, A. S., & ALLPORT, F. H. Personality structure and group structure: An interpretive study of their relationship through an event-structure hypothesis. *J. abnorm. soc. Psychol.*, 1956, **53**, 272–280.

THOMAS, E. J., & ZANDER, A. The relationship of goal structure to motivation under extreme conditions. *J. indiv. Psychol.*, 1959, **15**, 121–127.

TOLMAN, E. Principles of performance. *Psychol. Rev.*, 1955, **62**, 312–326.

ZIPP, SHEILA. An experimental study of resistance to influence. Unpublished doctoral dissertation. University of Michigan, 1958.

(Received August 3, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 110–118

RACE[1], SOCIAL CLASS, RELIGION, AND NATIONALITY AS DETERMINANTS OF SOCIAL DISTANCE[2]

HARRY C. TRIANDIS AND LEIGH MINTURN TRIANDIS

University of Illinois

SINCE Bogardus' (1928) pioneering work with an ordinal scale of social distance, in which a subject indicated zero social distance by stating that he was willing to marry a member of a particular ethnic group, and maximum social distance by stating that he would exclude such a person from the country, social distances scales have been used in a large number of studies. A reliable ranking of ethnic groups was made possible with data obtained with these scales. About 90% of white native-born Americans, it was found, are willing to marry Canadians, or Northern Europeans, but only 10% are willing to marry Southern and Eastern Europeans, and only 1% are willing to marry Negroes. About 10% of these subjects would exclude Jews, Serbo-Croatians, Negroes, Turks, and several other groups from the United States.

After Bogardus' use of the social distance scale with ethnic groups it was used by a number of researchers with other groups. For instance, Wilkinson (1929) used it with occupations, Bogardus (1933) with races, occupations, and religions. Sartain and Bell (1949) published equal appearing interval scale values. Melikian and Prothro (1952) obtained the social distance of Lebanese students towards national and religious groups and found greater social distance towards certain national groups than towards certain religious groups. The reverse result had been obtained by one of these authors in 1935. Prothro and Miles (1953) used the Sartain and Bell (1949) scale in the deep South. Ansari (1956) used the scale in India and found a relationship between social distance and stereotypes similar to that obtained by Katz and Braly (1933) in the United States. Hunt (1956) used the scale in the Philippines and found that his *Ss* showed little social distance towards Americans, Christians, and whites, but very much social distance towards Jewish, Moslem, and Buddhist groups. Kirsh (1957) used the scale in connection with voting behavior. Best and Sohner (1956) used it with political, economic, racial, religious, and nationality groups. They found very large social distances towards fascists, anarchists, and communists, large distances towards racial groups, small distances towards religious groups, and very small distances towards national groups. Hill (1953) has criticized the concept of social distance and presented arguments in favor of the usage of interviews and projective techniques for the determination of prejudice, rather than objective indices of social distance, such as the Bogardus scale.

The literature reviewed commonly recognizes that race, religion, nationality, occupation, or political ideology may be determinants of social distance, but does not explore how these factors combine and interact to determine the total social distance score. Furthermore, because of common ethnic stereotypes, these factors, when studied by the usual Bogardus technique, are confounded, e.g., since most Irishmen are Catholic, one cannot tell whether a reported prejudice towards the Irish represents prejudice towards the nationality or the religion; similarly, since Negro stereotypes imply lower class membership, class prejudice may contribute to reported Negro prejudice.

In addition, the literature suggests that some groups may stress *nationality* more than all other factors (e.g., the Lebanese in the 1950s), other groups *religion* (e.g., the Philippinoes), still other groups *race* (e.g., white Americans). Thus, the background characteristics of the subjects (*Ss*) may be related to the weight given to the various factors determining social distance.

The present study had three aims: first, to investigate the relative importance of race,

[1] Bogardus (1928) used the word *race* to describe the various ethnic groups judged by his subjects. In this paper we intend to use Hooton's well known definition of race as "a subdivision of mankind the members of which are distinguished by possession of similar combinations of anatomical characteristics due to common heredity."

[2] We are indebted to D. E. Dulany, L. G. Humphreys, W. J. McGuire, and I. D. Steiner for a critical reading of an earlier version of the manuscript and for valuable suggestions.

religion, nationality, and social class as determinants of social distance for our population of *Ss*. Second, to investigate the extent to which the relative importance of race, religion, nationality, and social class, as determinants of social distance, is dependent upon the subject's own background with respect to these factors. Third, to investigate the effect of the *S*'s race, religion, nationality, and social class upon his overall prejudice.

A factorial design, with two levels of race, two of social class, two of religion, and two levels of nationality was used. An analysis of variance of the social distance scores provided the information required by the first aim. Separate analyses of variance of the scores obtained from *Ss* of different backgrounds supplied information concerning the relative importance of the various factors for these different groups of *Ss*. The total variance of social distance scores, which is a measure of prejudice[3], was used in connection with the third aim. Finally, additional analyses, using protocols of individual *Ss*, investigated the effect of individual differences and the extent to which some *Ss* tend to use one factor, other *Ss* another, and still others various interactions between factors as the basis for their social distance scores.

METHOD

Standardization of the social distance scale. Thirty-six statements previously used by Sartain and Bell (1949) and 24 additional statements were given to 35 Illinois undergraduates who acted as judges. A graphic form of the Thurstone successive intervals procedure was used as described in Edwards (1957, Ch. 5). This procedure permits the development of an equal interval scale. The rank order correlation between the scale values obtained by this procedure and the scale values obtained by Sartain and Bell (using Texas *Ss*), for the 36 common statements, was .98. The statements that were finally selected for use are shown in Table 1. The data obtained from the 35 judges and the 15 statements that were finally selected were checked for unidimensionality, and a Guttman coefficient of reproducibility of .90 was obtained.

The scale values shown in Table 1 have been arranged so as to cover a 100-point scale with zero corresponding to the *S*'s endorsement of the statement "I would marry this person." This particular statement was not used in the study, however, because we wished to keep the sex of "this person" constant, and specified that it would be male; the statement was thus unsuitable for our male *Ss*. All other statements shown on Table 1 were used.

TABLE 1
STATEMENTS AND THEIR SUCCESSIVE INTERVAL VALUES ON A 100-POINT SCALE

1. I would marry this person.	0.00
2. I would accept this person as an intimate friend.	11.11
3. I would accept this person as a close kin by marriage.	21.50
4. I would accept this person as a roommate or I would date this person.	29.50
5. I would accept this person as a personal chum in my club.	31.10
6. I would accept this person as a neighbor	38.70
7. I would accept this person as my husband's or wife's friend.	40.90
8. I would live in the same apartment house with this person.	49.40
9. I would accept this person as one of my speaking acquaintances.	52.40
10. I would rent property from this person.	57.50
11. I would give asylum to this person, if he were a refugee, but I would not grant him citizenship.	63.10
12. I would not permit this person to live in my neighborhood.	69.70
13. I would not permit this person's attendance of our universities.	81.00
14. I would exclude this person from my country.	95.00
15. I would be willing to participate in the lynching of this person.	97.20

Subjects. *Ss* were 86 undergraduates from the introductory psychology class at the University of Illinois.

The questionnaire. An introduction was attached to the questionnaire in which it was stated that it is natural for people to "feel closer" to their relatives and friends than to strangers. *Ss* were urged to indicate how close they feel to various people described in the questionnaire. A statement to the effect that these people will often have unusual combinations of characteristics followed, in which certain extreme combinations, such as Portuguese Negro (from Portuguese Angola) were justified on the grounds that such people do in fact exist though the *S* may not be aware of it. *S* was then asked to indicate which religion among the Protestant, Catholic, Conservative or Reform Judaism varieties he considers *most different* from his own personal religious affiliation. He was asked to think of this particular religion when the words "different religion" appear as characteristics of the person described in the questionnaire. The questionnaire was anonymous, but information was obtained about the *S*'s sex, race, religion, ethnic background, and father's and mother's occupation and education.

A Public Opinion Questionnaire consisting of 14 statements[4] from the Adorno, Frenkel-Brunswik,

[3] The Total Sum of Squares (TSS) of an individual's social distance scores, towards the 16 stimulus people used in this study is a measure of prejudice, since the larger the TSS the greater the difference in the social distance experienced towards the various stimulus people. This may prove to be a particularly subtle way of measuring prejudice.

[4] The items are taken from Adorno et al. (1950), Pp. 255–257. They are No. 1, 4, 12, 13, 23, 26, 29, 34, 35, 37, 39, 42, 43, and 44.

TABLE 2
MEAN RESPONSES BY GROUP ON A 100-POINT SCALE OF SOCIAL DISTANCE

	Whites ($N = 77$)	Negroes ($N = 9$)
Swedish physician, white, same religion	11.6	23.0
Same religion, Greek physician, white	13.8	20.6
French physician, white, different religion	18.5	15.5
Banker, white, different religion, Italian	20.2	14.2
English coal miner, white, same religion	20.0	23.0
Greek truckdriver, white, same religion	21.0	17.0
Swede, truckdriver, white, different religion	23.2	19.4
White, different religion, Portuguese unskilled worker	32.1	23.0
French physician, Negro, same religion	34.2	11.0
Negro, same religion, Portuguese, civil Engineer	33.3	14.0
Civil engineer, Negro, different religion, French	35.4	14.0
Negro, different religion, Portuguese physician	36.8	19.5
Same religion, French, unskilled work, Negro	43.5	14.6
Negro, same religion, Portuguese, truckdriver	44.4	19.0
Different religion, French, unskilled work, Negro	44.6	19.0
Negro, different religion, Portuguese, coalminer	44.6	19.0

Levinson, & Sanford (1950) F Scale was placed at the end of the questionnaire. Since the present study makes only incidental use of the F Scale and the greatest number of published studies used the original scale rather than the modifications suggested by Sullivan and Adelson (1954), Bass (1956), and Jackson and Messick (1957), it was decided to use the scale in its original form.

The main body of the questionnaire consisted of the description of a person, e.g., "A white, Portuguese, unskilled worker, of the same religion as your own" and 15 statements from Table 1. The descriptions of the person are shown in Table 2. Four additional descriptions were used to assess reliability and are described below. *S* was asked if it was likely or unlikely that he would behave towards the person described (see Table 2) in the way described by the 15 statements of the social distance scale (see Table 1). The characteristics of the person were presented in a random sequence, as can be seen from Table 2. The 15 scale statements were also presented in a different random sequence on each sheet. Finally, the sheets were arranged, within the questionnaire, in different random sequences.

Determination of levels of race, social class, religion, and nationality. For levels of race we used white and Negro. For social class we used data from Welch (1949) and considered physician, banker, and civil engineer to be high, and unskilled worker, coalminer, and truckdriver to be low. For religion, we used the distinction between same religion and different religion. Finally, for nationality we employed the Bogardus (1928) data and considered the English, Swedes, and French high and the Portuguese, Italians, and Greeks as low.

Scoring. Since the unidimensionality of our social distance scale was adequate we scored it as follows: if only "positive statements" (scale value smaller than 58) were checked by *S* as being likely, we took as his score the most extreme positive statement checked as likely. If *S* checked only negative statements (more than 60) we took as his score the value of the most extreme negative statement that he considered as likely. When both positive and negative statements were endorsed, *S* received a score of 58. This method of scoring tended to increase the social distance score of many *S*s and produced a relatively more normal distribution of social distance scores. However, since most of our *S*s did not show much social distance the distribution of the scores was skewed, with a mode on the positive side of the neutral point.

Reliability. Fifty *S*s, taken at random from the 86, were used to compute reliabilities. The responses to the "French physician, white, and of different religion" were found to be correlated .96 to the responses to the "English banker, white, and different religion." The responses to the "Negro coalminer, same religion, Portuguese" and "Negro truckdriver, same religion, Portuguese" were correlated .90. Correlations of .69 and .87 were obtained for "French physician and civil engineer, Negro, same religion," and "White Portuguese and Greek unskilled workers" respectively. The average of these correlations is .855. When Fisher *Z*s are averaged, $r = .88$.

Classification of Ss. All *S*s were college students, thus, potentially middle class. In order to obtain some indications of how social distance is related to *S*s social class we categorized our *S*s as follows: Class I ("upper-middle"): At least one parent had a college education and a highly responsible occupation (e.g., physician, professor, executive in a large corporation). This corresponds very roughly to Warner's upper-middle or lower-upper class. Class III ("lower"): Neither parent had a college education and the father was a manual worker. This class corresponds roughly to Warner's upper-lower class. All other *S*s were classified in Class II ("lower-middle"). The *S*'s classification on ethnic background (high or low) was based on his answers to a question about his ethnic background. If he indicated old American, Northern or Northwestern European backgrounds, he was classified as High, otherwise as Low.

RESULTS

Relative Importance of Characteristics of Stimulus People

The first aim of this study was to investigate the relative importance of race, religion, nationality, and social class as determinants of social distance. Table 3 shows that race is the most important determinant of social distance for white *S*s, while interactions between factors are the important determinants

TABLE 3
ANALYSES OF VARIANCE OF SCORES OF 77 WHITES AND 9 NEGROES

Source	Whites				Negroes		
	df	*MS*	% Variance	*F*	*MS*	% Variance	*F*
Race	1	4.34	77.1	2170***	.10	15.4	—
Social Class	1	.93	16.6	465***	.10	15.4	—
Religion	1	.28	5.0	140***	.00	0.0	—
Nationality	1	.05	0.9	25***	.01	1.5	—
Interactions	11	.002	0.4		.042	67.7	—

*** p less than .001.

for Negro *Ss*. In other words, Negro *Ss* do not experience social distance towards whites, or persons with lower class occupations, or persons of different religion (e.g., Jews), but if the person in question has a combination of these "undesirable" characteristics, e.g., he is a "white, Jewish, Portuguese, unskilled worker," then Negro *Ss* feel some social distance as manifested in a refusal to accept such a person as a close kin by marriage, or as an intimate friend.

For white *Ss* race and social class are by far the most important determinants of social distance. They account for almost 95% of the variance of social distance scores. If we return to the previously used examples and consult Table 3, we can clarify the meaning of prejudice towards the Irish and the Negroes. Table 3 suggests that most of the prejudice towards the Irish is religious and not national; most of the prejudice towards Negroes is racial and not social class prejudice.

Background of Subjects and Factors Determining Social Distance

The second aim of this paper was to investigate the extent to which the relative importance of race, religion, nationality, and social class as determinants of social distance is dependent upon the *S*'s own background with respect to these factors. Table 2 shows differences between white and Negro *Ss* in mean social distance towards various stimulus persons. White *Ss* have a minimum of social distance towards the "Swedish physician of the same religion" and a maximum towards the "Negro, different religion, Portuguese, coalminer." Negro *Ss* show maximum social distance towards the "Swedish physician," the "English coalminer," and the "Portuguese unskilled worker," and minimum distance towards the "French physician, Negro, same religion." Table 3 shows a very striking difference between white and Negro *Ss* in the importance of the various factors as determinants of social distance. Additional analyses of variance, similar to the analyses presented in Table 3, were computed in which the *Ss* were grouped according to their social class, religion, ethnic background, sex, and F Scale scores.[5] These analyses indicate that the relatively less prejudiced groups (see below) emphasized race less than the more prejudiced groups. There was relatively little difference between the groups of *Ss* on the weight given to social class, religion, or nationality.

Background of Subjects as Related to Their Prejudice

The third aim of this paper was to investigate the effect of the *S*'s race, religion, nationality, and social class upon his overall prejudice. The Total Sum of Squares (TSS) of an individual's social distance scores, towards the 16 stimulus people, is a measure of prejudice (see Footnote 3). A comparison between the TSS of various groups is shown in Table 4.

Chi square tests computed from other data, obtained from the answers that the *Ss* gave on the instruction sheet of the questionnaire, indicate that Catholics consider the Jews rather than the Protestants as most different from themselves ($p < .02$), the Jews consider the Catholics as most different ($p < .01$), and the Protestants generally consider the Jews to be more different ($p < .20$). Almost a third of the Protestants, however, feel closer to the Jews than to the Catholics.

Catholics and conservative Jews have the highest F scores, when compared with the mean F score obtained from the Protestant group ($p < .01$; by chi square). Reform Jews have characteristically lower F scores than the Protestant average ($p < .01$; by chi square test).

[5] To save printing costs, these tables have been deposited with the American Documentation Institute. Order Document No. 6317 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

In an attempt to obtain a clearer picture concerning the significance of religion, social class, and national background, we considered the 18 possible subgroups that emerge when these three background factors are considered simultaneously. Tables 5 and 6 present the results of this analysis. The results obtained must be considered to be tentative since we have few cases in each cell. Some cells have only one *S*. The analysis of variance of Table 6 indicates that all three factors account for significant portions of the variance. National background is a particularly important background variable; yet it has not been sufficiently explored in studies of prejudice. The interaction between religion and national background also accounts for a good deal of the variance and nearly reaches the .05 level of significance. Most of the variance is accounted by national background, religion, and social class, in that order. This finding suggests that Westie and Westie (1957) have been working with a relatively less significant variable, social class, while religion, and national background are more important variables. A parsimonious interpretation of these data would, ideally, use a few constructs to account for all the findings, while the interpretations examined by Westie and Westie are attempting to explain only the relationship between social class and prejudice.

TABLE 4

COMPARISONS OF MEAN TOTAL SUMS OF SQUARES OF INDIVIDUALS IN VARIOUS GROUPS

Group	TSS	Group	TSS	Level of Significance (by *F* test)
Whites	5.62	Negroes	0.62	.001
High *F* scorers ($F > 3.5$)	9.32	Low *F*	4.71	.01
Lower class	6.56	Lower middle-class	5.90	.05
Lower class	6.56	Upper middle-class	4.30	.01
Lower middle-class	5.90	Upper middle-class	4.30	.05
Protestants	6.70	Jews	3.48	.01
Catholics	7.85	Jews	3.48	.01
Catholics	7.85	Protestants	6.70	.05
Women	7.16	Men	5.15	.01
High Ethnic background *Ss*	7.24	Low Ethnic background *Ss*	3.92	.01

Notes.—All groups consist of whites only, except for the first row.

Analysis of Individual Protocols

In addition to the group analyses of variance, we also analyzed individual protocols. The social distance scores of some *Ss* appear to be determined mostly by one factor, and those of some other *Ss* by another factor. For some *Ss* combinations of factors, as evidenced by significant interactions in the analysis of variance of their scores, seem more important than any single factor. These analyses also showed that some *Ss* deviate from the general pattern (greatest importance to race, then social class, then religion, and minimal importance to nationality) and show no social distance because of race, but a great deal of social distance because of occupation.

TABLE 5

AVERAGE TOTAL SUM OF SQUARES OF SOCIAL DISTANCE SCORES OF *Ss*, CLASSIFIED ON THE BASIS OF THE *S*'S BACKGROUND

Social Class	Catholics	Protestants	Jewish
Upper-middle	High Ethnic = 9.15 ($n = 3$) Low Ethnic = 4.95 ($n = 1$)	High Ethnic = 5.75 ($n = 7$) Low Ethnic = 3.10 ($n = 2$)	High Ethnic = 4.24 ($n = 1$) Low Ethnic = 2.29 ($n = 5$)
Lower-middle	High Ethnic = 14.50 ($n = 4$) Low Ethnic = 7.85 ($n = 3$)	High Ethnic = 7.12 ($n = 10$) Low Ethnic = 3.84 ($n = 1$)	High Ethnic = 5.25 ($n = 7$) Low Ethnic = 2.84 ($n = 7$)
Upper-lower	High Ethnic = 10.00 ($n = 3$) Low Ethnic = 5.40 ($n = 8$)	High Ethnic = 9.70 ($n = 10$) Low Ethnic = 5.25 ($n = 2$)	High Ethnic = 7.90 ($n = 1$) Low Ethnic = 4.26 ($n = 4$)

Note.—High Ethnic = *Ss* whose parents come from N or NW Europe; Low Ethnic = *Ss* whose parents come from S or E Europe; only white *Ss* were used for this table.

TABLE 6
ANALYSIS OF VARIANCE OF DATA OF *S*'S BACKGROUND (Table 5)

Source	*df*	*MS*	*F*	% Variance
Social Class (Sc)	2	14.35	8.7*	16.6
Religion (R)	2	27.20	16.5*	31.4
National Background (N)	1	40.50	24.6**	23.4
Sc × R	4	4.74	2.87	1.1
Sc × N	2	1.40	—	1.6
R × N	2	10.55	6.4	12.2
Error	4	1.65	6.4	3.8
Total	17			

* $p < .05$.
** $p < .01$.

DISCUSSION

For the *S*s of this study race is by far the most important factor determining social distance. Furthermore, social class is a much more significant determinant of social distance than is religion or nationality.[6] Social class is relatively more important for groups that show little social distance and have low F scores than for groups showing a great deal of social distance ($p < .05$). The absolute amount of variance due to social class, as distinguished from the percentage of variance due to social class, is about equal in all our groups. That is, the groups showing greater total social distance do so because they discriminate much more on the basis of race than do other groups.

None of the group data show any interaction effects. On the other hand, the analyses of individual protocol show that several interactions determine significant portions of an individual's social distance. It appears that there is no systematic effect of any particular interaction, over all *S*s, so that the individual interaction effects are eliminated when the group score analyses are computed.

Interpretations of the phenomena of social distance and prejudice have often emphasized psychodynamic factors (e.g., Adorno et al., 1950). Allport (1954) has presented a more adequate analysis employing a multilevel approach in which historical, sociocultural, and situational variables interact with psychodynamic variables and the stimulus itself. On the other hand, Allport does not specify which of these factors is most important. The present data are clearly apposite.

Several of our findings do not require interpretation. That white *S*s show more social distance than Negro *S*s, that high F score *S*s show more distance than low F score *S*s is not surprising. Since all of our stimulus people were male, the greater social distance shown by women may be an artifact of the testing procedure, since it may include the social distance between the sexes.

The remaining differences, however, need some interpretation. Briefly, we find that Catholics are more prejudiced that Protestants, and Protestants more prejudiced than Jews. The middle classes are less prejudiced than the lower class; finally, *S*s of lower ethnic background are less prejudiced than those of higher ethnic background.

Our attempt to interpret these data uses three basic constructs: conformity, cognitive dissonance, and insecurity. Conformity concerns the adoption of the values of the ingroup. This is unquestionably an extremely significant phenomenon; it has been established by a large number of studies, the best known of which is probably Newcomb's (1943). The fact that national background is so important a background variable suggests that conformity may be by far the most important determinant of social distance. Allport (1954) is probably right when he concludes that "about half of all prejudiced attitudes are based only on the need to conform to custom, to let well enough alone, to maintain the cultural pattern" (p. 286).

The second construct, cognitive dissonance, is adopted from Festinger (1957). In applying dissonance theory to prejudice, we take economic competition, with the socioeconomic inequalities that result from it, as the starting point of prejudice. Prejudice is a device that permits justification of the inequalities that result from economic competition. The person who first started competing for something, in the dawn of history, was not prejudiced. However, after establishing an advantage over other human beings, he felt the need to justify

[6] These findings must be interpreted as referring only to the average range of variables, as they appear in the United States today. If the range of some of the variables is increased, for example, if pagans, atheists etc., were introduced, a different apportionment of the variance would be obtained.

it by arguing that his opponents had deserved to lose because of their inherent inferiority. More explicitly, in terms of Festinger's theory, we find on the one hand the existing social and economic inequalities which result in behaviors that Dollard (1957) has described involving economic, sexual and prestige gains; on the other hand there is the American creed, as described by Myrdal (1944). Cognitive dissonance arises when a person behaves in ways that are incompatible with the American Creed ("all men are created equal . . . "), the Christian ethic or other systems of ethics implying equality in the eyes of a supernatural being or the state. Dissonance is decreased when the "exploited" group is regarded as inferior, since the democratic ideal does not then apply to them. With respect to an aspect of anti-Semitism, dissonance may operate in the opposite direction. The Jew is more often in middle-class than in lower-class occupations. This tendency towards upper middle-class occupations may make the Jew appear to be doing "too well", which may lead to the stereotype that he is pushy, shrewd, grasping, intelligent, sly, and aggressive.

The dissonance theory of prejudice, that we have presented so far, does not account for the fact that within a group of "exploiters", some become bigots, but others become tolerant, and still others militant liberals. It seems that here the psychodynamic factors described by Adorno et al. (1950) become important. The authoritarian personality is characteristic of a person who is threatened, *insecure* (Gough, 1951), highly conservative, rigid and intolerant of ambiguity (Frenkel-Brunswik, 1949). The authoritarian appears to have a need to reduce the ambiguity involved in high dissonance to a much greater extent than the tolerant individual. The insecure person cannot tolerate dissonance, whereas the secure person can tolerate it.

Adorno et al. (1950, p. 385) found that child training practices were related to the development of authoritarian personalities. *Ss* scoring high on the F scale reported that their home discipline was harsh, threatening, and arbitrary. Hart (1957) also found that prejudiced *Ss* have been subjected to harsh and threatening discipline. Bronfenbrenner's (1958) analysis of the trends in socialization practices in the United States clearly states that

> in matters of discipline working-class parents are consistently likely to employ physical punishment, while middle-class families rely more on reasoning, isolation, appeals to guilt, and other methods involving the threat of loss of love (p. 424).

Also,

> Over the entire 25-year period studied, parent-child relationships in the middle class are consistently reported as more acceptant and equalitarian, while those in the working class are oriented towards maintaining order and obedience" (p. 425).

To sum up the theoretical argument, child training practices may be responsible for the development of a relatively secure, or a relatively insecure personality. The insecure person is usually incapable of tolerating ambiguity. Since in most highly developed societies there are alternative sets of values, which are often contradictory (e.g., equalitarian values may contradict the value that the in-group is better than the out-group), the insecure person attempts to reduce the dissonance generated by such contradictions by fully adopting one of the available ideologies.

With this theoretical scheme in mind we may re-examine the data. First, we find that lower-class *Ss* are more prejudiced than middle class subjects. The child training practices used by the working classes are those which are likely to produce insecurity. This is not true for the child training practices of middle-class parents. This finding, then is consistent with the theory that insecure persons are more prejudiced than secure person.

The differences found to be associated with social class are, in turn, consistent with the findings of other investigators. For instance, Westie and Westie (1957) find maximum social distance between lower-class whites and Negroes, and minimal social distance between upper-class whites and Negroes, with middle-class whites falling between lower- and upper-class whites on their social distance scores.

Second, we find greater prejudice in our Catholic than in our Jewish *Ss*, with the Protestants in between. Psathas (1957) gave a questionnaire to high school students with Southern Italian and Eastern European Jewish backgrounds, concerning the child training practices of their parents. The one significant difference between the responses of these two ethnic groups was that the Jewish students reported that their parents had a much greater

regard for the children's ability to think for themselves. This stress on intellectual independence extended to all areas of thought and Jewish children were taught "to question authority in (their) schooling since even Divine Law is subject to interpretation" (Psathas, 1957, p. 415). This type of socialization is obviously less authoritarian than that of the Catholics. It trains children to tolerate ambiguity, since values are in a state of flux and subject to interpretation. O'Rielly and O'Rielly (1954) found that Catholic *Ss* scoring high on a scale of intensity of religious belief were less favorable to Jews and Negroes and favored segregation of Negroes in their own parishes. It would seem, then, that certain features of the socialization practices of Catholic parents, particularly the clarity of their approach towards the Divine Law, predispose their children to intolerance of ambiguity, which in turn is correlated with prejudice. Jews, on the other hand, seem to train their children in a manner which minimizes these tendencies.

The third finding of the present study was that *Ss* with Old American, Northern or Northwestern European ethnic backgrounds show more social distance than *Ss* with Southern or Eastern European backgrounds. In fact, national background was next to religion our most important background variable. An examination of the analyses of variances of the scores of these two kinds of *Ss* reveals that the difference between the two groups with respect to social distance is entirely accounted by the difference in the amount of distance that they experience toward Negroes. This finding cannot be fully interpreted, in the absence of evidence on the child training practices of these groups. Conformity to established cultural norms may be an important factor accounting for it.

The foregoing analysis appears to account for our data, for anti-Negro and anti-Semitic prejudice, with a minimum of theoretical constructs. Like most simple theories it is probably too simple. It seems, however, a reasonably parsimonious first step toward a theory of prejudice. It is limited to an explanation of prejudiced attitudes and does not include an explanation of prejudiced behavior. For such an explanation an extension of the model is required, which in our view should bring considerations from frustration-aggression theory, recently reviewed by Cowen, Lander, & Schaet, (1959), to bear on the probability of occurrence of prejudiced behavior. The theory also provides a series of testable hypotheses about the relationship between prejudice and child training practices, social differentiation, and value conflict.

SUMMARY

In a study of social distance of college students with respect to various social objects, a factorial design with two levels of value of race, social class, religion, and nationality was employed and analyses of variance were computed on social distance scores. For white *Ss* race and social class were found to be more important determinants of social distance than religion or nationality. The relative importance of the various factors, however, varied from group to group. White *Ss* were found to show more social distance than Negro *Ss*; high F score *Ss* more than low F score *Ss*, women more than men, Catholics more than Protestants, and these more than Jews, lower-class *Ss* more than lower middle-class *Ss* and these more than upper middle-class *Ss*, and *Ss* from Northern and Northeastern Europe more than *Ss* with ethnic backgrounds from Southern and Eastern Europe. Analysis of the protocols of individual *Ss* showed that there is a tendency for the more prejudiced *Ss* to assign more of the variance of social distance scores to race than to social class. The data are interpreted in terms of a theory of prejudice that employs conformity, cognitive dissonance, and insecurity as its main constructs.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

ALLPORT, G. W. *The nature of prejudice*. Cambridge, Mass: Addison-Wesley, 1954.

ANSARI, ANWAR. A study of the relation between group stereotypes and social distance. *J. educ. Psychol.*, Baroda, 1956, **14**, 28–35.

BASS, B. M. Development and evaluation of a scale for measuring social acquiescence. *J. abnorm. soc. Psychol.*, 1956, **53**, 296–299.

BEST, W. H., & SOHNER, C. P. Social distance and politics: Some selected relationships. *Sociol. soc. Res.*, 1956, **40**, 339–342.

BOGARDUS, E. S. *Immigration and race attitudes*. Boston: Heath, 1928.

BOGARDUS, E. S. A social distance scale. *Sociol. soc. Res.*, 1933, **17**, 265–271.

BRONFENBRENNER, U. Socialization and social class through time and space. In Eleanor E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1958. Pp. 400–425.

COWEN E., LANDER, J., & SCHAET, D. E. The effects of mild frustration on the expression of prejudice attitudes. *J. abnorm. soc. Psychol.*, 1959, **58**, 33–38.

DOLLARD, J. *Caste and class in a southern town*. New York: Anchor, 1957.

EDWARDS, A. L. *Techniques of attitude scale construction*. New York: Appleton-Century-Crofts, 1957.

FESTINGER, L. *A theory of cognitive dissonance*. Evanston, Ill: Row, Peterson, 1957.

FRENKEL-BRUNSWIK, ELSE. Intolerance of ambiguity as an emotional and perceptual personality variable. In J. S. Bruner & D. Kretch (Eds.), *Perception and personality: A symposium*. Durham, N. C.: Duke Univer. Press. 1949. Pp. 108–144.

GOUGH, H. G. Studies in social intolerance: A personality scale for antisemitism. *J. soc. Psychol.*, 1951, **33**, 247–255.

HART, I. Maternal child-rearing practices and authoritarian ideology. *J. abnorm. soc. Psychol.*, 1957, **55**, 232–237.

HILL, M. Some problems of social distance in intergroup relations. In M. Sherif & M. O. Wilson (Eds.), *Group relations at the crossroads*. New York: Harper, 1953. Pp. 285–307.

HUNT, C. L. Social distance in the Phillippines. *Sociol. soc. Res.* 1956, **40**, 253–260.

JACKSON, D. N., & MESSICK, S. J. A note on "ethnocentrism" and acquiescent response sets. *J. abnorm. soc. Psychol.*, 1957, **54**, 132–134.

KATZ, D., & BRALY, K. W. Verbal stereotypes and racial prejudice. In Eleanor E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1958. Pp. 40–46.

KIRSCH, A. D. Social distance and voting behavior related to N variables. *Dissertation Abstr.*, 1957, **17**, 2673.

MELIKIAN, L., & PROTHRO, E. T. Social distance and social change in the Near East. *Sociol. soc. Res.*, 1952, **37**, 3–11.

MYRDAL, G. *An American dilemma: The Negro problem and modern democracy*. New York: Harper, 1944.

NEWCOMB, T. M. *Personality and social change*. New York: Dryden, 1943.

O'REILLY, C. T., & O'REILLY, E. J. Religious beliefs of Catholic college students and their attitudes toward minorities. *J. abnorm. soc. Psychol.*, 1954, **49**, 378–380.

PROTHRO, E. T., & MILES, O. K. Social distance in the deep South as measured by a revised Bogardus scale. *J. soc. Psychol.*, 1953, **37**, 171–194.

PSATHAS, G. Ethnicity, social class and adolescent independence from parental control. *Amer. sociol. Rev.*, 1957, **22**, 415–423.

SARTAIN, A. Q., & BELL, H. V., JR. An evaluation of the Bogardus scale of social distance by the method of equal appearing intervals. *J. soc. Psychol.*, 1949, **29**, 85–91.

SULLIVAN, P. L., & ADELSON, J. Ethnocentrism and misanthropy. *J. abnorm. soc. Psychol.*, 1954, **47**, 246–250.

WILKINSON, F. Social distance between occupations. *Sociol. soc. Res.* 1929, **13**, 234–244.

WELCH, M. K. The ranking of occupations on the basis of social status. *Occupations*, 1949, **27**, 237–241.

WESTIE, F. R., & WESTIE, MARGARET L. The social distance pyramid: Relationships between caste and class. *Amer. J. Sociol.*, 1957, **63**, 190–196.

(Received July 21, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 119–123

OCCUPATIONAL CHOICE AND COGNITIVE FUNCTIONING:

CAREER ASPIRATIONS OF HIGHLY INTELLIGENT AND OF HIGHLY CREATIVE ADOLESCENTS[1]

J. W. GETZELS AND P. W. JACKSON
University of Chicago

THE RELATIONSHIP between a person's cognitive style and his orientation to the world is of central importance to the understanding of both cognitive functioning and social behavior. The present study investigates the relationship between cognition, as defined by performance on intellective tasks, and personal orientation, as defined by long-term occupational choice and career aspiration.

The study is part of a research program examining the fantasy productions, school performance, and teacher preferences of two types of adolescents (*a*) those exceptionally high in intelligence but not equally high in creativity, and (*b*) those exceptionally high in creativity but not equally high in intelligence. Results already reported indicated: (*a*) despite striking differences in IQ, the two groups were equally superior in school achievement to the population from which they were drawn; (*b*) despite this equal school achievement, the high IQ students rather than the high Creativity students were preferred by teachers; and (*c*) there were significant differences in the fantasy productions of the two groups, the creative adolescents expressing more stimulus-free, imaginative, humorous, and playful themes.[2] The present investigation examines the relationship between the career choices and the cognitive functioning of the two groups.

The purpose of the investigation is two-fold: At the substantive level, to determine the occupational preferences of the two groups; at the theoretical level, to ascertain the relationship between types of cognition and types of career aspiration. Two guiding empirical questions were posed: (*a*) what is the nature of the long-term occupational goals of the two groups? (*b*) to what extent do perceived "teacher standards" and the criterion "success in adult life" enter as motives in the aspirations of the two groups?

METHOD

The population from which the experimental groups were drawn comprised 449 adolescents (largely from academic and professional families) who made up the total student body of a Midwestern private secondary school. The groups were chosen on the basis of their performance on the following tests:

Standard IQ tests

Either a Binet, a WISC, or a Henmon-Nelson score was available for each adolescent. The scores obtained from the WISC and the Henmon-Nelson were converted by regression equation to comparable Binet IQs.

Five creativity measures

These were taken or adapted from Guilford and Cattell, or constructed especially for the study as follows:

1. *Word Association.* The *S* was required to give as many definitions as possible to fairly common stimulus-words, e.g., BOLT, BARK, etc. His score depended upon the absolute number of different categories into which these definitions could be put.

2. *Uses for Things.* The *S* was required to give as many uses as he could for objects that customarily have a stereotyped function attached to them, e.g., BRICK, PAPER-CLIP, etc. His score depended upon both the number and originality of the uses which he mentioned.

3. *Hidden Shapes.* The *S* was required to find a given geometric form that was hidden in more complex geometric forms or patterns.

4. *Fables.* The *S* was presented with four fables in which the last lines were missing. He was required to compose three different endings for each fable: a moralistic, a humorous, and a sad ending. His score depended upon the number, appropriateness, and originality of the endings.

5. *Make-Up Problems.* The *S* was presented with

[1] The research described in this report was supported by a grant from the United States Office of Education. The present report is an expanded version of a paper read at the APA meeting, Cincinnati, September 3, 1959.

[2] J. W. Getzels & P. W. Jackson, "The Highly Creative and the Highly Intelligent Adolescent: An Attempt at Differentiation," Paper read at the APA meetings, Washington, D. C., September, 1958; "The Social Context of Giftedness: A Multidimensional Approach to Definition and Method," Paper read at the American Sociological meetings, Seattle, Washington, August, 1958; "The Meaning of Giftedness: An Examination of an Expanding Concept," *Phi Delta Kappan*, November, 1958, 75–78.

four complex paragraphs, each of which contained several numerical statements, e.g., "the costs involved in building a house." He was required to devise as many mathematical problems as he could that might be solved with the information given. His score depended upon the number, appropriateness, and originality of the problems.[2]

On the basis of the IQ measures and a summated score on the five creativity instruments, the two experimental groups were formed as follows:

1. Students exceptionally high in IQ but not equally high on the creativity measures (hereafter referred to as the High Intelligence Group). These students were in the top 20% in IQ when compared with like-sexed age peers, but *below* the top 20% on the creativity measures. Their mean IQ was 150, with a range from 139 to 179. $N = 28$ (17 boys, 11 girls).

2. Students exceptionally high on the creativity measures but not equally high in IQ (hereafter referred to as the High Creativity Group). These students were in the top 20% on the creativity measures when compared with like-sexed peers, but *below* the top 20% in IQ. Their mean IQ was 127, with a range from 108 to 138. $N = 26$ (15 boys, 11 girls).[4]

Four instruments were used to obtain the data with respect to the two experimental questions:

1. *A Direct Sentence Completion Test.* Among the 60 items comprising this instrument were such career preference items as:

When I grow up I want to become a(n) . . . .

If the choice of a permanent occupation were left up to me, I would choose . . . .

The work that has the greatest appeal to me is . . .

I always wanted to be a(n) . . . .

The *S* was required to respond "directly" to these items by completing each stub so as to "reflect his own preferences." The responses of the two groups were analyzed for the number and kind of occupations mentioned.

2. *An Indirect Sentence Completion Test.* This instrument was parallel to the Direct test, except that the sentence stubs were written in the third person using boys' names for the male *S*s and girls' names for the female *S*s:

When Stan (Jean) grows up, he (she) wants to become a(n) . . . .

If the choice of an occupation were left up to him (her), Ken (Kate) would choose . . . . (etc.)

The *S* was required to respond "projectively" to the items by completing the stubs as a "verbal speed test." The responses of the two groups were analyzed for the number and kind of occupations mentioned.

3. *A Parent Questionnaire.* This instrument required the parents to respond to two questions regarding their children's occupational goals as follows:

Has the pupil expressed interest in a particular career? If yes, specify . . . .

The responses from the parents of the two groups were analyzed for the number and kind of occupations mentioned.

4. *An Outstanding Traits Test.* This instrument contained descriptions of 13 children, each of the 13 exemplifying some desirable personal quality or trait. For example, one child is described as having the highest IQ in the school, another as being the best athlete, another as having the best sense of humor, another as being the best-looking, another as being the most creative person in the school, and so on. The *S*s were required to rank these 13 hypothetical children in three ways: (*a*) on the degree to which they would like to be like them, (*b*) on the degree to which they believe teachers would like them, and (*c*) on the degree to which they believed people with these various qualities would succeed in adult life. The rankings given by the two groups within each category and the relationships between the rankings for the different categories were analyzed.

RESULTS

The findings may be discussed with respect to each of the two questions in turn:

1. Differences between the two groups appear both in the *quantity* and *quality* of occupational goals. The *quantity* of occupational possibilities mentioned is significantly greater for the high Creatives than for the high IQs, the latter group mentioning an average of 1.82 different occupations on the Direct Sentence Completion Test and 3.57 on the Indirect Test, the former group mentioning an average of 2.61 and 5.00 on the respective instruments. The *quality* of the different occupations mentioned as possibilities is also significantly different for the two groups. When the occupations mentioned are analyzed into conventional (lawyer, doctor, professor) and unconventional (adventurer, inventor, writer) categories, it is found that the high Creatives give a significantly greater proportion of unconventional occupations than do the high IQs. For example, on the Direct Sentence Completion Test, 16% of the high IQs mention an unconventional occupation, 62% of the high Creatives mention at least one such occupation as a possibility. The complete data are presented in Table 1.[5]

[2] The descriptive data with respect to these tests will appear in the technical report being prepared for the United States Office of Education. In general, the intercorrelations among these instruments for the total population ranged from .079 to .466, with a mean of approximately .30.

[4] The mean IQ for the total student body was 132, with a standard deviation of 15.1.

[5] It is interesting to note that mention of unusual occupations by the high Creatives occurs with equal frequency in their "direct" and their "projective" responses, while for the high IQ's mention of unusual occupations is much more likely to occur in their "projective" responses than in their "direct" responses. This difference supports the proposition held by Kubie

TABLE 1

QUANTITY AND QUALITY OF OCCUPATIONS MENTIONED BY THE EXPERIMENTAL GROUPS ON DIRECT AND INDIRECT SENTENCE COMPLETION TESTS

Test	Group	Number of Occupations Mentioned			Unusual Occupations	
		Total	$\bar{X}$	σ	Number Mentioned	Number of Ss Mentioning
Direct	IQ ($N = 28$)	51	1.82*	1.09	6	5**
	Creative ($N = 26$)	68	2.61*	1.41	24	16**
Indirect	IQ ($N = 28$)	100	3.57**	1.81	12	10*
	Creative ($N = 26$)	130	5.00**	1.80	29	17*

t used to test differences between means in Col. 2.
χ^2 used to test differences between frequencies in Col. 5.
* Significant at .05 level.
** Significant at .01 level.

The preceding findings based on the reports of the adolescents themselves are amply supported by the data from the Parent Questionnaire. Parents of 24 of the high IQ children and of 24 of the high Creativity children responded to the Questionnaire. Sixteen of the high IQ parents and 20 of the high Creativity parents reported that their children had expressed an interest in some career. What is most noteworthy in these reports is that 12 or 75% of the high IQ parents mention occupations falling into but *five* career categories: engineering and architecture, science, medicine, law, and teaching; only 7 or 35% of the high Creativity parents mention career choices falling exclusively into these five categories ($\chi^2 = 5.71$, $p < .02$). Instead, the high Creativity parents tend to mention such rather unconventional occupations as "veterinary" or "entertainer," such unresolved combinations as "law or music," "teaching or art," and such expressive professions as "writing" or "dancing," none of these latter being mentioned as possibilities by the parents of the high IQ children.

In short, with respect to the *quantity* of occupations considered, depending on one's point of view, one may say that the high Creativity adolescents as compared with the high IQ adolescents are more "diffuse" in occupational goals or that they are "more able and willing to deal with a greater range of career possibilities." With respect to the *quality* of occupations considered, again depending on one's point of view, one may say that the high Creativity adolescents are more "eccentric" in their occupational goals, or that they are "more able and willing to deal with career risks, that is, to take greater liberties with accepted standards of adult success."

2. Data with respect to the second empirical question, i.e., the "success" orientation of the two groups, were obtained from the Outstanding Traits Test, and provide further insight into the differential career aspirations of the high IQ and high Creativity adolescents.

The high IQs ranked the qualities in which they would like to be outstanding in the following order:[6] (1) character, (2) emotional stability, (3) goal directedness, (4) creativity, (5) wide range of interests, (6) high marks, (7) IQ, (8) sense of humor. The high Creatives ranked the qualities in the following order: (1) emotional stability, (2) sense of humor, (3) character, (4.5) wide range of interests, (4.5) goal-directedness, (6) creativity, (7) high marks, (8) IQ. Most noteworthy here is the extraordinarily high ranking given by the high Creativity group to "sense of humor," a ranking that not only distinguishes them from the high IQ group (who gave it a rank of 8) but also from the student body from which the experimental groups were drawn (who gave it a rank of 7.5).

The most relevant and striking differences, however, between the two groups in the context of the present inquiry are observed in the relationship between the qualities they want for themselves, the qualities they believe lead to adult success, and the qualities they believe teachers favor. For the high IQ group, the

and others pointing to the "flux" or movement of preconscious material into the conscious of the creative person, and the relative restriction of such movement in the more highly intellectualized person. Follow-up of these suggestive findings is anticipated.

[6] The entire population of the school almost without exception ranked social skills first as the quality in which they would like to be outstanding, and athletics, good-looks, high energy level, and health last. In view of this very high agreement, the responses were re-ranked omitting the uniformly-ranked qualities. The findings reported here are on the re-ranked data.

TABLE 2
RANK ORDER CORRELATIONS AMONG SUBSECTIONS OF THE OUTSTANDING TRAITS TEST

Components of Correlation	Subjects	
	IQ (N = 28)	Creative (N = 26)
Personal traits believed predictive of success and Personal traits believed favored by teachers	.62	.59
Personal traits preferred for oneself and Personal traits believed predictive of adult success	.81	.10
Personal traits preferred for oneself and Personal traits believed favored by teachers	.67	−.25

rank order correlation between the qualities they would like to have themselves and the qualities they believe make for success in adult life is .81; for the high Creativity group, it is .10. For the high IQ group, the rank order correlation between the qualities they would like to have themselves and the qualities they believe teachers favor is .67; for the high Creativity group, it is (minus) −.25. The data are presented in Table 2.

In effect, the high IQ adolescent wants the qualities he believes make for adult success and the qualities that are similar to those he believes teachers like, the high Creative adolescent favors personal qualities having no relationship to those he believes make for adult success and are in some ways the reverse of those he believes his teachers favor. Again, depending upon one's point of view, one may say that the high Creativity adolescents as compared with the high IQ adolescents are more "rebellious" or more "autonomous" with respect to adult standards of success; or conversely, that the high IQ adolescents are more "compliant" or more "realistic." But in any case, what seems clear is that there is a systematic relationship between types of cognitive functioning and types of career aspiration. The high IQ and the high Creativity groups are characterized by differences in their career aspirations—differences all the more *basic* since the school achievement of the two groups as measured by standardized verbal and numerical subject-matter tests was the *same*.

DISCUSSION

Several existing conceptual formulations may be adduced to account for the present data. In the context of cognitive theory, for example, Guilford's (1957) factors of *convergent* and *divergent* thinking are highly relevant. Discussing the production of tests to assess these factors, he states:

> In tests of convergent thinking there is almost always one conclusion or answer that is regarded as unique, and thinking is to be channeled or controlled in the direction of that answer .... In divergent thinking, on the other hand, there is much searching about or going off in various directions .... Divergent thinking ... is characterized ... as being less goal-bound. There is freedom to go off in different directions ... Rejecting the old solution and striking out in some new direction is necessary, and the resourceful organism will probably succeed. (pp. 6, 7, 9.)

It is of course clear that in these terms the high IQ adolescents tend to favor "convergent" modes of problem-solving, and the high Creativity adolescents tend to favor "divergent" modes of problem-solving—whether in the field of cognitive functioning as represented by performance on the intelligence and creativity measures, or in the field of personal-social functioning as represented by occupational and career choice.

In the context of motivational theory, Maslow's (1956) formulations of Defense and Growth is similarly relevant to the present issues. He writes:

> Every human being has both sets of forces within him. One set clings to safety and defensiveness out of fear, tending to regress, hanging on to the past ... afraid to take chances, afraid to jeopardize what he already has, afraid of independence, freedom, separation. The other set of forces impels him forward toward wholeness of Self and uniqueness of Self, toward full functioning of all his capacities, toward confidence in the face of the external world at the same time that he can accept his deepest, real unconscious Self.... This basic dilemma or conflict between the defensive forces and the growth trends I conceive to be existential, imbedded in the deepest nature of the human being, now and forever into the future....
>
> Therefore we can consider the process of healthy growth to be a never-ending series of free choice situations, confronting each individual at every point throughout his life, in which he must choose between the delights of safety and growth, dependence and independence... Safety has both anxieties and delights; growth has both anxieties and delights. (pp. 37–38.)

Whether we assume the cognitive or the motivational stance, adopt for our use the one set of concepts or the other, or whether we attempt what appears to us the more fruitful procedure of seeing the two formulations as referring to similar processes in merely different terms, it seems to us that the essence of the performance of our high Creativity adolescents might lie in their ability to produce new forms, to risk conjoining elements that are customarily thought of as independent and dissimilar. In contrast, the high IQ adolescent seemed to possess to a high degree the ability, and perhaps the need, to focus on the usual, to be "channeled and controlled" in the direction of the right answers, the socially accepted solution.

Moreover, these differences between the high IQs and the high Creatives are not restricted to their intellectual performance alone or to their occupational choice alone. The data with respect to both the cognitively-oriented and the socially-oriented behavior are of a piece, and the characteristics that describe the one describe the other. The high IQs seem to converge upon stereotyped meanings, to perceive personal success by conventional standards, to move toward the model provided by teachers, to seek out careers that conform to what is expected of them. The high Creatives seem to diverge from stereotyped meanings, to move away from the model provided by teachers, to seek out careers that do not conform to what is expected of them. In fine, it seems to us that the outstanding feature of all the data is the consistency of the cognitive (as defined by performance on intellective tasks) and the personal-social (as defined by occupational choice and career aspiration) aspects of behavior.

SUMMARY

This investigation extends an earlier series of studies of two types of gifted adolescents, those exceptionally high in IQ but not in creativity and those exceptionally high in creativity but not in IQ, by comparing their career aspirations and their cognitive functioning. It was found that (*a*) the two groups differ significantly in both the quantity and quality of occupational goals, the high Creatives giving a greater number of occupations and more "unusual" occupations, and (*b*) the two groups differ significantly in attitudes toward adult success and toward conformity to teacher preferences, the high Creatives being less concerned with conventional standards in both areas. These relationships between "intellective" and "personal-social" behavior are discussed in terms of Guilford's cognitive model of "convergence" and "divergence" in intellectual performance and Maslow's motivational model of "safety" and "growth" in social performance.

REFERENCES

GUILFORD, J. P. *A revised structure of intellect.* Rep. Psychol. Lab., No. 19, Los Angeles, Univer. Southern California, April 1957.

MASLOW, A. H. Defense and growth. *Merrill-Palmer Quart.*, 1956, **3**, No. 1, 35–47.

(Received August 3, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 124–137

SUBLIMINAL PERCEPTION OR PERCEPTION OF PARTIAL CUES[1]

MORTON WIENER AND PETER H. SCHILLER
Clark University

IF THE number of investigations is a valid criterion of importance of an area, "perception without awareness" or "subliminal perception" has been of considerable theoretical as well as of practical concern (e.g., Adams, 1957; Goldiamond, 1958; Jenkin, 1957; McConnell, Cutler, & McNeil, 1958). Interest in this area seems to reflect the influence of personality theories that postulate an ability of the human organism to make discriminating responses to stimuli of which the organism is "unaware" but which are important in determining the behavior of the organism; and the influence of theories that attribute a generally influential role to personality in the determination of perceptual processes. Much experimentation has been initiated with many ingenious approaches, but the findings and conclusions still appear to be equivocal.

Examination and analysis of this vast body of theoretical and empirical data suggests that two basic explanatory models have evolved to account for the phenomena in question (i.e., "perceptual defense," "need perception" and "perception without awareness," etc.). One type of explanation, referred to here for purposes of exposition as the "two-process" view, posits two distinct perceptual processes, a supraliminal one and a subliminal one (e.g., subception). These two processes are integrated, but also operate relatively independently, and different responses may be evoked by each. In this two-process model, a critical stimulus when not perceived via the supraliminal process (consciously) may be perceived via the subliminal process (not consciously); the subliminal process may then set off the appropriate need-related or defense-related responses[2] (Blum, 1955; Klein, 1956; Lazarus & McCleary, 1951; McGinnies, 1949).

The second explanatory model, referred to here as the "one-process" view, posits a single perceptual process that accounts for the phenomena in question as well as for perceptual behavior in general. The one-process view maintains that "perception without awareness" is essentially a misnomer, insofar as total unawareness of the stimulus in subliminal research has not been adequately determined. Rather, the advocates of the one-process view claim that the experiments cited as demonstrating subliminal processes (e.g., Lazarus & McCleary, 1951) from which subliminal processes have been inferred have not adequately investigated the possibility that some supraliminal cues were present. These theorists would hold that it is the perception of these cues that sets off the so-called "need-related" or "defense-related" responses; "need-related" or "defense-related" responses are conceptualized as response hierarchies or response probabilities. Whether one infers "defense" or "sensitization", for example, is a function of a match between the response hierarchy of the

[1] The three studies reported here were supported in part by a United States Public Health Research Grant M-2568A to the senior author. Experiment I is based on a thesis completed by Peter H. Schiller under the direction of the senior author and submitted in partial fulfillment of Peter H. Schiller's requirement for an MA degree at Clark University. Portions of Experiment I were presented at the Eastern Psychological Association Meetings, Atlantic City, April 1959. Portions of Experiment II were presented at the Eastern Psychological Meeting, New York, April 1960. The authors wish to thank William Vogel for his help in collecting the data; Eugene Sacks, Kenneth Wagoner, Donald MacLean, Harold Teegarden, Donald Theobald, for help in early pilot work for Experiment II. Experiment III is also part of a series of studies now being carried out by Mr. Schiller designed to try to separate the "difficulty" factor of the intentional tasks from the amount of the intentional task learned and their effects upon incidental learning. We would also like to thank Joachim Wohlwill and Seymour Wapner for their very helpful comments.

[2] The theoretical view that there is "inhibition" of the perception of the noxious stimulus in "perception without awareness" is somewhat more difficult to categorize in the two basic positions proposed here. On one hand, the inhibition model usually posits a single perceptual process; yet on the other, it apparently also posits a "gating" process for channeling the input. In some manner, particular inputs are selectively channeled out, i.e., inhibited from reaching consciousness but are responded to at some other level. In the broadest sense, this would seem to be categorized best with the "two-process" view. (Bruner, 1957; Perloe 1959).

S and the presented stimulus (Goldiamond, 1958; Wiener, 1957).

An elaboration of some of the methodological and logical issues may help clarify the position of each of the two theoretical models. "Perception without awareness" is inferred by those who hold the two-process view whenever the following three conditions occur: (*a*) a stimulus is presented at some intensity (or duration, or clarity) below the level at which *S*s could correctly identify the stimulus on 100% of the trials; (*b*) the *S* cannot or does not correctly identify the stimulus verbally or motorically (consciously); (*c*) some response occurs (e.g., autonomic, or imagery) that would be expected to occur only if the stimulus were somehow perceived and identified.

In psychophysics, "threshold" and "subliminal" are arbitrary statistical concepts. Threshold is usually defined as some intensity of stimulation (or duration) where the *S* can correctly identify the stimulus (or its presence) a certain percentage of the time, usually 50%.

Thus, in psychophysical research when the threshold point is identified, it is usually determined as the point where the *S* *correctly* identifies the stimulus 50% of the time and does *not correctly* identify the stimulus 50% of the time, over a group of trials. Subliminal designates any point of stimulus presentation below this arbitrary point (e.g., from 0% through 49% correct).

Further, because a homogeneous patch of light has usually been the stimulus for establishing the threshold in vision research, it has been acceptable to infer that no portion of the stimulus is perceived if the *S* is unable to identify its presence correctly. However, when a complex stimulus (i.e., a word, or a picture, or a design) is used and presented at a statistically determined threshold point, there is some question whether it is equally reasonable in this instance to assume that *no portion* of the presented stimulus has been perceived, even when the response is incorrect. For when a complex stimulus is presented at any specified threshold intensity, theoretically at least, any one of the following perceptual experiences could occur on *any* single trial: (*a*) the presented stimulus is supraliminally perceived, and the stimulus is correctly identified either by verbal or other motor response; (*b*) some portion of the stimulus is supraliminally perceived, and the stimulus is correctly identified; (*c*) some portion of the stimulus is supraliminally perceived, and the stimulus is incorrectly identified; (*d*) the stimulus is not supraliminally perceived, nor is any portion, and the stimulus is correctly identified (e.g., by chance); (*e*) the stimulus is not supraliminally perceived, nor is any portion, and the stimulus is incorrectly identified. It is important then, in this area of research, to differentiate clearly whether subliminal implies only an intensity of stimulus presentation that results in some percentage of correct report of the stimulus less than 50%, and incorrect identification of the stimulus, or *no* awareness at all by the *S* of the stimulus or of any portion of it.

At issue, then, is whether perception without awareness has been inferred from the occurrence of both *c* and *e* alternatives above, or only on the basis of Alternative *e*. The one-process view assumes that the experimental demonstrations from which "perception without awareness" has been inferred by the two-process view are due to the effects elicited by occurrence of Alternative *c* above; the two-process view assumes that the occurrence of *e* underlies the conclusion drawn with respect to "perception without awareness."

However, despite the possibility of fluctuation of threshold and variations in threshold level for different figures or words, most experimenters have failed to control or assess the threshold level for *each* trial, and therefore the exact perceptual conditions for every trial is uncertain. The one-process theorists maintain that the *S*, on some of the "incorrect" trials at least, perceives some portion of the stimulus, perhaps not sufficient to make a correct report, but that the *S* makes a response which is some portion of the total correct response he would have made if he had perceived all of the stimulus (Ericksen, 1956), and it is these responses to partial cues, that have been attributed to the subliminal process. Some evidence has been cited to support this latter conclusion. It has been demonstrated (Bricker & Chapanis, 1953; Murdock, 1954) that although the first guesses of *S*s are incorrect, under tachistoscopic presentation, correct second guesses are significantly better than could be expected by chance. These investi-

gators conclude that the *S*s must have seen something, although not enough to lead to a correct verbal guess. Voor (1956) found that the subception effect decreased as less information (i.e., lower brightness of stimulus) was available to the *S*.

Thus, to hold the view that subliminal effects occur (two-process model) requires an assertion of the negative proposition, there is *no* evidence of supraliminal effects that can account for the noted behavior. To hold the one-process view requires that evidence be presented that supraliminal cues are present in these instances. It would, therefore, seem the responsibility of those who hold the one-process view to demonstrate that supraliminal effects *are* present and can account for the findings that have been interpreted as evidence of subliminal processes; and the responsibility of those who hold the two-process view to rule out, as carefully as possible, any plausible supraliminal cues, when trying to demonstrate subliminal processes.[3]

In the research to be reported, a series of experiments was undertaken, each a paradigm of the type usually used to demonstrate subliminal perception, to explore whether the effects that apparently occur can be used as a basis for inferring a subliminal process, or whether the effects can be accounted for by the alternative view, i.e., perception of partial cues.

These studies attempt to explore carefully the perceptual conditions in these paradigmatic experiments to determine whether partial cues are perceived, or whether it is reasonable to conclude that since there is no evidence of any supraliminal cue perception, subliminal perceptual processes can be inferred.

EXPERIMENT I: SUBLIMINAL STIMULATION IN THE CHOICE BEHAVIOR PARADIGM

In this paradigm, the *S* is generally presented with a stimulus at some intensity below the point at which 100% correct report occurs; most often this point is set at a very low percentage threshold point. The point typically is determined by a series of trials at the beginning of the experiment. However, since thresholds fluctuate for *S*s with practice, fatigue, motivation; and since any of several perceptual possibilities may occur on *any single trial* at any threshold level (as noted above), it is essential that the exact conditions of awareness of the stimulus be investigated for every trial. Unless this control is established, we cannot conclude that the noted effects of change in choice behavior (or problem solving) under low percentage threshold intensity are attributable to trials on which no portion of the stimulus is supraliminally perceived (rather than the possibility that the effects are attributable to trials on which some portion of the stimulus is supraliminally perceived, although incorrect identification occurs).

The purpose of this experiment is to investigate each of the above possible alternatives by systematically exploring changes in choice behavior when a stimulus is presented at a low percentage threshold intensity. Are the changes in choice behavior a function of those trials in which no supraliminal cues are evident, or can the effects be attributed to trials in which supraliminal cues can be demonstrated?

In this experiment, if the effects are attributable to trials on which all or part of the stimulus is supraliminally perceived, it will not be possible to conclude that *no* effects are ever attributable to trials on which *no* portion of the stimulus is supraliminally perceived—a null hypothesis conclusion. Rather, it would be evidence in support of the one-process view which holds that overall effects which have been in the past cited as evidence for subliminal process effects must be questioned, and that demonstration of subliminal effects will have to be shown to occur when there is *no* apparent evidence of perception of partial cue correlated with the noted effects, for each trial.

Method

Apparatus and procedure. Ten ambiguous line drawings, similar to the Carmichael, Hogan and Walters (1932) series, were presented on a screen by means of a projector. Beneath each design were two words, either of which could be the name for the ambiguous design. The task for the *S* was to choose the word that seemed more applicable. Wiener, Carpenter, & Carpenter (1957) have shown that the choice behavior of a *S* can be modified by this procedure.

In order to control word order choice effects, two sets of slides were prepared. The second set was identical to the first in all respects, except for the reversal of the

[3] It is apparent that "subliminal perception" or "subception" cannot be "proven", i.e., it is not possible to prove *no* supraliminal effects are present; nor can the presence of positive evidence of supraliminal processes *disprove* the two-process view either, i.e., there are *no* effects due to a subliminal process.

two words below the figures. On the screen, the figures and the words appeared white on a black background.

Two projectors were employed in the study. The first was used to expose the ambiguous designs and the words beneath them. Both designs and words were exposed well above absolute threshold. The second was used to project the subliminal stimulus, a circle, so that it surrounded one or the other word appearing underneath the ambiguous designs. For different experimental groups the circle was projected either with a constant exposure or with intermittent exposure to investigate whether the effects were different for these two types of stimulation. For the constant presentation a conventional 100 w. projector lamp was employed; the power supply to this lamp was put through a transformer. By varying the transformer output and by use of filters, the intensity level of the projector light could be varied. For the intermittent presentation a tachistocopic neon light was placed into the light compartment of the projector. The light was operated by means of a timing unit which was triggered off three times per sec. by a synchronous motor and microswitch.

The screen was 15′3″ from the projector. In the experimental room four desk chairs, each 9′3″ from the screen, were secured to the floor in a semicircle. A 25-w. shielded light was placed 11 ft. above the floor, over the desks. This light stayed on throughout the experiment.

Each ambiguous design was presented for a period of 15 sec., followed by a 15-sec. interval. The *S*s recorded their choices after each presentation, and then checked one of the following three statements for each choice:

(*a*) definitely saw a circle
(*b*) possibly saw a circle but not sure
(*c*) definitely did not see a circle

Before the appearance of each slide a buzzer was sounded to alert the *S*s to the next exposure.

The instructions to the *S*s were as follows:

> In this experiment I am going to show you some slides. Each slide will show a picture of an object. Below the object you will see two words, like you see on this card in my hand (sample card shown). You are to indicate which of the two words is correct for the picture shown on the slide. This experiment was performed before, under slightly different conditions, on a large group of students. They were shown the same pictures, but for a full minute. For each picture that group decided which one of the labels was correct. This time we will show you the same pictures for a somewhat shorter period, and we want to see which label you select. In order to help you to make a choice, sometimes the answers found correct by the other group will be circled, like this [sample card shown]. In some cases the circle will be shown quite clearly. In other cases not so clearly, and still other times the circle will not be shown at all. If you do not see the circle, you will have to make your decision without considering the other group's choice. I will give each of you a sheet of paper now. As you can see, after each number there is a line. After a slide has been shown, write the answer you think is correct on this line. Also, next to the line there are three statements. Please circle the one that applies to you in each case. Please watch the screen during the entire time the picture is shown. Write down your answer and check the appropriate statement only after the picture has been removed. The appearance of each new slide will be indicated by the sound of a buzzer. Do you have any questions? Remember, whenever you see a circle around one of the words, the circle means that that word was chosen by our other group as the correct label. I will turn the light off now. Above you there is a little light which is just bright enough for you to see by. Please go ahead now and fill out the two lines on top of your sheet.

Subjects. 144 undergraduate college students were divided into six groups: two control and four experimental groups. Each of the six groups were further subdivided into two groups: one was shown the ambiguous designs with the choice words in one order; the other subgroup had the order of choice words reversed.

The six groups were as follows:

Control Group I—The circle was never presented.

Control Group II—The circle was presented during each 15 sec. exposure at an intensity where it was plainly visible.

Experimental Group I—The circle was presented constantly during each 15 sec. exposure at a level (52 v. together with a filtering lens) where a group of sophisticated judges had correctly guessed the location of the circle 95% of the time.

Experimental Group II—The circle was presented constantly during each 15 sec. exposure at a brighter intensity (64 v.).

Experimental Group III—The circle was projected intermittently three times per sec., with each duration lasting .07 sec. This level was determined by another group of sophisticated judges (as in Experimental Group I).

Experimental Group IV—The circle was projected intermittently three times per sec. with each duration lasting .09 sec.

For every group, with the exception of Control Group I, the same word was circled for each stimulus design. The word circled for each design was the one less frequently selected by Control Group I. Since this group was not shown a circle at any time, their choices represented the influence of the ambiguous designs alone.

Results

The score for each *S* was the number of the circled words chosen by that *S* as the appropriate label for the designs. Two types of analyses were performed. In the first, the data for the six groups were treated by analysis of variance and by *t* tests. In the second analysis, all of the choices for the *S*s in the experimental group were analyzed according to the subjective reports for each trial. These data were further analyzed by chi square analyses for each design.

Table 1 indicates that there are statistically significant differences among the groups ($p < .01$) in the number of circled choices chosen. Other comparisons show no statistically significant differences.

Table 2 shows the mean number of circled words chosen by each group, the variance, and the percentage of each type of verbal report. As the percentage of the reports of seeing the circle increased, there was an increase in the mean number of circled choices (the circle was always presented below the absolute threshold level for the experimental group *Ss*).

Table 3 shows that there are significant differences between the two control groups as well as among those groups differing considerably in the number of subjective reports of the amount of times the circle was seen. No significant differences are found between Control Group 1 (no circle presented) and those experimental groups who had the circle presented at the lowest levels of presentation, either constantly or intermittently.

To assess the influence of the experimental stimulus (the subliminally presented circle) under each of the three subjective reports, all the choices for the *Ss* from all of the experimental groups were combined. The number of circled words chosen under each of the three subjective report categories was determined. Table 4 shows the percentage of choices of the circled words for each of the subjective categories and the two control groups. The Not Seen category results in choice behavior very similar to the Control Group I (not shown a circle), while the other two categories are similar to the Control Group II (always shown a circle).[4]

The results of the analyses by chi square show that for 6 of the 10 designs the number of choices for each design is significantly different ($p < .05$) between the two Control Groups. Choices for 4 of the 10 designs are significantly different between the category Not Seen and the combined Definitely Seen and Possibly Seen. There were no significant differences found in the analyses between the Control Group I and the Not Seen category.

TABLE 1

ANALYSIS OF VARIANCE OF THE NUMBER OF CIRCLED CHOICES SELECTED BY EACH OF THE SIX GROUPS WITH TWO ORDERS OF PRESENTATION

Source	*DF*	*MS*	*F*
Between Groups	5	18.4	5.09*
Between Orders	1	1	—
Interaction	5	3.8	1.05
Error	132	3.6	

* $p < .01$.

TABLE 2

MEAN NUMBER OF CIRCLED WORDS CHOSEN, VARIANCE, AND PERCENTAGE OF SUBJECTIVE REPORTS

Group	Mean	Variance	Definitely Seen (%)	Possibly Seen (%)	Not Seen (%)
Control Group I	3.33	2.68	.4	.4	99.2
Control Group II	5.62	4.66	99.9	.1	0
Experimental Group I	3.62	2.60	2.5	9.2	88.3
Experimental Group II	4.62	5.48	49.1	16.2	34.7
Experimental Group III	3.96	1.93	16.6	23.7	59.7
Experimental Group IV	5.00	4.27	30.8	17.0	52.2

TABLE 3

T TEST ANALYSIS BETWEEN GROUPS OF THE MEAN NUMBER OF CIRCLED WORDS SELECTED BY EACH GROUP

	Exp. I	Exp. II	Exp. III	Exp. IV	Control II
Control I	.617	2.349*	1.147	3.041**	4.171**
Experimental I		1.821*	.615	2.513**	3.642**
Experimental II			1.202	.692	1.821*
Experimental III				1.894*	3.023**
Experimental IV					1.129

* $p < .05$.
** $p < .01$.

TABLE 4

PERCENTAGE OF CIRCLED WORDS CHOSEN BY THE CONTROL GROUPS AND UNDER THE THREE SUBJECTIVE REPORT CONDITIONS

Group	Per cent
Control Group I	33.3
Not Seen Group	36.5
Possibly Seen Group	53.5
Definitely Seen Group	52.5
Control Group II	55.8

[4] In this instance, partial information is apparently equally effective in influencing behavior. This finding is not surprising; uncertain information or partial information would be expected to communicate as much information for choice decision—in a two-choice situation—as total information.

In addition the overall magnitude of these latter chi squares is also the smallest. Four of the 10 chi square analyses are significantly different between the Control Group I and the Possibly Seen category.

In summary the analyses of the data yield the following results:

1. Comparison of Experimental Group I, II, III, and IV with Control Group I would appear to demonstrate that subliminal effects have occurred.

2. However, the mean number of circled words chosen increased significantly as exposure brightness increased.

3. Analysis of the data according to subjective reports shows a significantly greater percentage of choices of circled words on those trials on which the *Ss* reported seeing or possibly seeing the circle, than on those for which they reported not seeing it. Under the Definitely Seen *and* the Possibly Seen categories the number of choices approximated the choices of Control Group II (always shown the circle). When the *Ss* reported not seeing the circle although it had actually been presented, their choices were not different from the Control Group I, i.e., no circle presented at all.

4. There were no significant differences noted between constant and intermittent presentations of the subliminal stimulus.

Discussion

The results of this study support the hypothesis of "perception of partial cues." The experimental stimulus, i.e., the circle presented below absolute threshold, appeared to be effective in influencing choice behavior only when the subliminally presented influencing stimulus was reported Seen or Possibly Seen. In most experiments, since the circle would not be identified in the Possibly Seen conditions, these would have been considered true subliminal trials. However, there was apparently no effect upon choice behavior when the *Ss* reported not seeing the stimulus at all. Thus, under the conditions of this experiment, choice behavior was modified by subliminal presentation (i.e., the circle presented below the absolute threshold) only on those trials in which there was partial or complete verbal awareness of the experimental stimulus. These results do not require the inference of subliminal processes.

The question can be raised whether Possibly Seen is equivalent to "perception of partial cues", particularly when a simple stimulus is presented (e.g. a circle). This question requires some additional research. However, it does not seem unreasonable to assume that uncertainty is based in part at least upon some perception of partial cue. Whether or not these are equivalent, the effects noted in other studies that used this type of stimulation procedure and in which *no* perception was inferred because the stimulus was not or could not be identified, must be more carefully assessed to determine whether the noted effects were attributable to the uncertainty (perception of partial cue?) trials rather than to the Not Seen trials.

EXPERIMENT II: SUBCEPTION OR PERCEPTION OF PARTIAL CUE IN THE "AUTONOMIC DISCRIMINATION WITHOUT AWARENESS" PARADIGM

One assumption in the two-process view of need-perception and defense-perception is that discrimination of stimuli (particularly need and defense stimuli) can take place subliminally, and that such discrimination is quite similar in effect to the discriminations that take place in the supraliminal process. Although there are suggestions that the subliminal process may function somewhat differently than the supraliminal, the two-process view has always implied that the subliminal process can and does distinguish stimuli, and evoke responses appropriate to the need-related, or the defense-related stimuli. Further, this discrimination must be based primarily on the "meaning" of the stimulus. Unless stimulus words were subliminally identified and categorized in terms of "need" or "threat", how could the organism respond to STEAK when hungry, or avoid words like WHORE if repressed for sexual stimuli? Without this ability to identify and categorize somehow a correct response could not be expected. The alternative would be to assume there are special structural characteristics of neutral, need, or defense stimuli, a not very tenable assumption. Whatever the process by which subliminal discrimination occurs, the end result is that the discrimination and the

responses follow along dimensions that are similar to those which are usually attributed to supraliminal cognitive processes. If, then, the subliminal process evokes responses to stimuli that belong to the same class (i.e., WHORE or PROSTITUTE) and does not evoke this same response to neutral words (i.e., WHERE or CONSTITUTE), nonclass stimuli, then it is possible to investigate the one- and two-process view by a careful analysis of autonomic generalization behavior, in the "subception" paradigm.

Lazarus & McCleary (1951) apparently demonstrated "subception"—a process by which an autonomic discrimination is made to stimuli presented below absolute threshold, even when the *S* is unable to make a correct supraliminal discrimination. Other investigators (Rubenfeld, Lowenfeld, & Guthrie 1956) showed that the subception effect involves stimulus generalization.

Riess (1940) has demonstrated that when verbal stimuli are presented supraliminally the galvanic skin response (GSR) generalizes more to words that are semantically related to the originally conditioned word than to words structurally similar to the supraliminally conditioned words, but there is some generalization to both.

The purpose of this experiment is to investigate the GSR generalization gradients to words semantically similar to the conditioned word (e.g., ship–boat) as compared with words structurally similar (e.g., ship–shop) under conditions of low percentage threshold presentation and incorrect identification of the stimulus.

If it is maintained that the organism "discriminates" a subliminal stimulus and responds to it as belonging to a threat or shock category and the response is greater than to those stimuli which belong to a nonthreat or nonshock category (autonomic discrimination without awareness), then words that mean the same should also show such response. On the other hand, if it is maintained that the effects attributed to subception are attributable to perception of partial cue, (e.g., Bricker & Chapanis, 1953; Voor, 1956) then it is expected that the GSR response should occur to structurally similar words rather than to the words that have the same meaning but are structurally dissimilar. Further, the amount of response to the subliminally presented stimuli should be expected to be a decreasing function as the amount of stimulus perceived by the subject decreases.

Method

Material. The Galvanic Skin Response was continuously measured and recorded with a Grass Model #5 Polygraph. The electrodes were two penny-size sterling silver discs, coated with silver chloride and embedded in foam rubber and strapped to the *S*'s palm.

Shock was administered from 1½ v. dry cells through a Student's Inductorium to electrodes strapped to the leg of the *S*.

The tachistoscope was an electronically triggered, two-field apparatus, built by Roy Gunter Associates of Clark University.

The stimulus words were printed on a tape and presented tachistoscopically.

All of the recordings of GSR and presentations of the stimuli were automatic. A 2 cpm motor and a series of microswitches made it possible to (*a*) present an alerting signal (a red light), (*b*) trigger the tachistoscope at the manually set exposure level, (*c*) mark the time of the stimulus presentation on the recording tape, (*d*) administer shock if scheduled, (*e*) present a verbal response signal (green light), after which the *S* rested until the next alerting signal, and (*f*) move the next stimulus into the tachistoscopic field. This cycle was completed every 30 sec. The shock was administered approximately two sec. following the presentation of the conditioned stimulus.[3] The response signal was presented six sec. after the onset of the conditioned stimulus.

The stimulus words (printed in capitals) were:

Shock	*Semantically Similar*	*Structurally Similar*	*Nonshock*
START	BEGIN	STATE	COULD
SHIP	BOAT	SHOP	EDGE
CENTER	MIDDLE	RECENT	SECOND
ALLOW	PERMIT	FOLLOW	LAUGH
NEARLY	ALMOST	MERELY	DURING

These words were carefully matched for frequency of usage in the Thorndike-Lorge word count. Pilot work indicated there were no apparent differences in thresholds for any category, nor for any particular word or words.

Subjects. The *S*s were all volunteers who were members of a college fraternity. In return for the *S*s' participation, the fraternity was paid for the time they spent in the experiment. In this way, the *S*s felt they

[3] Although the literature (White & Schlosberg, 1952; Wickens, 1959) suggests that maximal conditioning occurs if the unconditioned stimuli follow the conditioned stimulus by approximately .5 sec. Pilot work indicated that there appeared to be rapid extinction of the conditioned response under a low level aperiodic reinforcement schedule on those trials on which there were no reinforcement with .5 sec. delay interval.

were contributing their efforts to research, yet they recognized they were receiving some payment for the discomfort of participating in a shock experiment. Each *S* had the option of having the shock level lowered if he felt it was too painful, and of stopping at any point in the experiment. Three *S*s did not complete all phases of the experiment.

Twenty-three *S*s completed all portions of the study, but only 12 records met the final three criteria established for testing the hypothesis.

Procedure. To enhance the possibility of conditioning, the *E* was curt, somewhat gruff and impersonal from the time the *S* entered the experimental room until the *S* was finished. It has been discovered in pilot work that the more anxious the *S*, the more easily he seemed to condition to the shock. The *S* was seated in a comfortable chair. Shock apparatus was then attached to his left leg. The GSR recording electrodes were attached to the *S*'s left palm. The *S* was seated behind a cardboard screen. He could see no equipment except the eyepiece of the tachistoscope. A clip board hung near the eyepiece.

Conditioning phase. A card with 10 words, 5 of them with asterisks, was hung on the clip board. The *S* was given the following instructions:

> If you look into the T scope you will see a green glow. When that green glow changes to a red glow, it means, pay attention, for a word will flash on the screen you see back here in the tachistoscope. You are to wait until the red glow changes to green, and then you are to tell me what the word was. In other words, when the red glow flashes, pay attention; when the green glow comes back on, report what the word was and relax. No word will appear in the T scope that does not appear on this card. When a word marked on this card with an asterisk appears, you may or may not be shocked; at first you will be shocked every time a word with an asterisk on this list appears, after that, you may or may not be shocked when a word with an asterisk appears. You will never under any conditions be shocked when a word appears which is not marked on the card with an asterisk. Understand? Let's go.

Conditioning was on a 100% reinforcement schedule for two presentations of each shock word; thereafter, it was an aperiodic 50% schedule. After four presentations of all of the five shock and the five nonshock words, a test trial was given. Just prior to this test trial, the *S* was asked if he could recall the five shock words. A shock was administered as soon as the *S* recalled all five shock words. The criterion of conditioning was that each shock word elicited a higher GSR than each nonshock word on the test trial. If this criterion was not met, the procedure was continued with intermittent test trials following further conditioning trials. In this conditioning phase, words were presented at a supraliminal level, although an attempt was made to keep the exposure time relatively short. The conditioning period lasted between 40 min. and 2 hr.

Subliminal test phase. After a rest period of 5 to 15 min., the *S* was given the following instructions:

> Now I am going to present the words at very fast speeds. You are to tell me what the word was when the green light comes back on. You may not see the word every time; nevertheless *guess*. That's important. If you're not sure, *guess*. Guess from this card [a new card was tacked up on the clip board] There are 20 words here, 10 of them are new; the other 10 you've already seen. You will be shocked, if you will be shocked, only on those words to which you were shocked before. By the way, you haven't been reacting too well; I'm going to raise the shock.

The shock was not in fact raised unless the *S* had shown by his GSR reactions that he had adapted to the shock intensity. Each of the 20 words was presented five times at a point as close to the *S*'s 50% level of recognition as possible. The approximate exposure level was determined by early pilot work. This level was adjusted for each *S* individually as the experiment progressed. Whenever the *S* had a run of trials on which he correctly identified the stimulus, the stimulus exposure duration was decreased; when the *S* had a run of incorrect identifications, the duration was increased. These changes were necessary to keep close to the 40–60% threshold point. The exposure level for all classes of words was not different, since the words from each class were randomly ordered in the list. An aperiodic shock schedule was administered through this phase. There were two shock trials for every five exposures of the shock words.

Supraliminal test phase. During this phase *S* was simply told, "Now I'll show you words at slower speeds, so you can recognize each one." Although no shock was given, the *S* was not so informed and each *S* appeared to expect to be shocked. Each of the 20 words was shown once at a clearly supraliminal level.

Results

Verbal responses. For the 23 *S*s who completed all phases of the experiment, it was possible to examine the incorrect verbal guesses for the subliminal test phase trials. For each shock word (conditioned stimulus), there were two generalization words, one meaning, one structure. Incorrect verbal guesses could be analyzed for the mean number of times a shock word was incorrectly guessed as being its matched meaning word (e.g., SHIP presented—guessed BOAT) .30; or as its matched structure word (e.g., SHIP presented—guessed SHOP) 1.57; and for the number of times the meaning words were incorrectly guessed as its shock pair (e.g., PERMIT presented—guessed ALLOW) .39; and the number of times the structure words were incorrectly guessed as its shock pair (e.g., FOLLOW presented—guessed ALLOW) 2.70. Chance is .30. These differences are reliable. Twenty-one *S*s have more matched pair shock-structure and structure-shock guesses compared to their meaning-shock and shock-meaning guesses. The other two *S*s had

equal numbers of incorrectly guessed pairs. Each of these two *S*s had two pairs in the two categories of incorrect guessing.

Autonomic responses. Three criteria were established to determine whether or not each *S*'s record would be analyzed to test the hypothesis. First, over all trials the stimulus words should have been presented between the 35–65% threshold level. One *S* who correctly identified 71% of the stimulus words was excluded as not meeting this criterion. The second criterion was that in the supraliminal test phase, the GSR for at least four of the shock words had to be greater than the GSR to *any* of the nonshock words. Seven of the 23 *S*s did not meet this criterion. The third criterion was that the mean GSR evoked by the shock words had to be greater than the mean GSR evoked by the nonshock words for those trials of the low percentage threshold presentation when the words were correctly identified by the *S*. Three *S*s were excluded as not meeting this criterion. Earlier investigators suggested that these three criteria were minimal requirements to test the hypothesis of this study. Voor (1956), for example, had noted that the conditioning seemed to disappear when the transition was made from relatively slow exposure speeds of the conditioning trial to the relatively fast exposure speeds of the stimulus presentations in the low percentage threshold level test phase. It was evident that if the generalization hypothesis were to be tested, it would be necessary to insure that conditioning was present for each *S* throughout the experiment. The latter two criteria imply that at least minimal conditioning was present for each *S* included in this analysis. Twelve *S*s remained who met all three criteria.

A Percentage Change Score was computed for each stimulus presentation. Three measures of resistance were determined, one at 1 sec., one at 2 sec. and one at 3 sec. prior to the presentation of the stimulus. Each of these readings was converted to conductance units $\left(\frac{1}{R}\right)$. The base level score then for each trial was the average of these three readings. The Percentage Change Score was determined by taking the maximum deflection reading within the 7 sec. following the stimulus presentation (usually between the 4th and 5th sec.), converting this reading to conductance units, and subtracting the computed base level score from this value and then dividing this difference by the base level and multiplying by 10^4 to remove decimals.[6] If the maximum deflection after stimulus presentation was less than the average base level score, the Percentage Change Score was estimated as zero for that trial.

TABLE 5

ANALYSIS OF VARIANCE OF THE PERCENTAGE CHANGE SCORE FOR THE THREE PRESENTATION CONDITIONS AND FOR THE FOUR CLASSES OF STIMULI

Source	*DF*	*MS*	*F*
Between Conditions (C)	2	11,512	3.06
Between Stimuli (S)	3	56,811	3.31*
Between Individuals (I)	11	184,566	9.48**
C × S	6	19,461	9.64***
I × C	22	17,178	8.51***
I × S	33	3,761	1.86*
I × C × S	66	2,019	

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Each *S* was then given an average Percentage Change Score for each of the four classes of stimulus words (i.e., shock, meaning, structure, neutral) for his Supraliminal Trials, for those low percentage threshold presentation trials on which he correctly identified the stimulus word, and for those low duration presentation trials on which he did *not* correctly identify the stimulus word. Thus each *S* had 12 scores. These scores were analyzed by an Analysis of Variance. Because the interactions were significant, the first order effects were tested against the interactions. Table 5 indicates that all effects were significant, except Between Conditions, (i.e., Supraliminal, Correct Subliminal, and Incorrect Subliminal trials). However, the Conditions by Stimulus interaction was statistically significant ($p < .001$). Examination of Fig. 1 shows that the three other classes of stimuli elicited greater responses than did the neutral words, and that

[6] Percentage Change Score =

$$\frac{\text{Maximum conductance score} - \text{Base level conductance score}}{\text{Base level conductance score}} \times 10^4$$

the response to the meaning words is not less than to the structure words on the Supraliminal trials. The effect changes on the correct identification trials of the Subliminal presentation. On the trials where the *Ss* incorrectly identify the words under low duration presentations, the response to the structure words is greater than all of the other three categories. Noteworthy, too, is the successive decrease in the amount of response to the shock words as the amount of identification decreased (and inferentially the amount of stimulus the *Ss* saw of the word).

It has been noted that the GSR is a function of the verbal response, as well as the response to the stimulus. To be certain that these effects cannot be attributed to different guessing responses to structure words than to meaning words in those trials in which stimuli presented subliminally were incorrectly identified, a further analysis was performed on the scores for each *S* for each stimulus category (i.e., shock, meaning, structure, neutral words) categorized by whether the *S* had guessed it as a shock or as a nonshock word. Because six of the 12 *Ss* did not have entries in the cell for shock word presented, incorrect shock word guessed, this category and that for shock word presented, nonshock word guessed incorrectly were excluded from the analysis. Three *Ss* failed to have a single cell entry in one of the six remaining cells for each *S*. These values were estimated by successive approximations, taking into account both row and column values. The three estimated values were entered in the analysis of variance, and the degrees of freedom for the analysis were reduced by three. Table 6 shows the results of this analysis of variance. Since none of the interactions was significantly different from zero, the sums of squares and degrees of freedom were combined into a pooled error term. Significant *F* values were found between stimuli ($p < .05$) and between individuals ($p < .001$). The variances attributed to kind of guesses (shock word or nonshock word) did not quite reach a significant level (*F* value 3.06; 3.29 needed for $p = .05$). Examination of the means of the mean GSR to the stimulus words indicated that the noted significant differences among the means can be attributed to the higher GSR evoked by the structure words than by the other two categories. Further examination showed that the amount of the *Ss*' responses to the structure words was greater for both kinds of guesses—incorrect guessed shock and the incorrect guessed nonshock—although this difference is much greater in the former instance. Figure 2 shows these relationships. It appears clear that the effects attributed to subliminal processes appear to be a function of the GSR responses to the incorrectly identified structure word presentations.

TABLE 6

ANALYSIS OF VARIANCE OF THE PERCENTAGE CHANGE SCORE OF THE THREE STIMULUS CLASS WORDS GUESSED SHOCK AND NONSHOCK FOR THE INCORRECT SUBLIMINAL TRIALS

Source	*DF*	*MS*	*F*
Between Guesses (G)	1	6,347	1.37
Between Stimulus Words (S)	2	18,686	4.03*
Between Individuals (I)	11	55,943	12.05**
G × S	2	13,935	2.46
G × I	11	2,218	—
S × I	22	4,128	—
Residual	19	5,662	
Pooled Error	54	4,642	

* $p < .05$.
** $p < .001$.

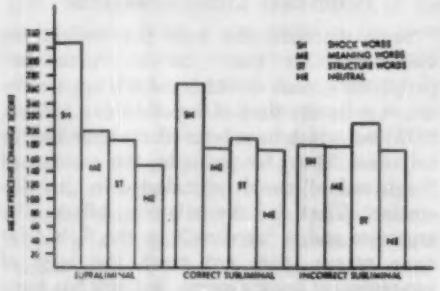

FIG. 1. Mean per cent change score of GSR for each stimulus category for the three stimulus conditions.

Discussion

These results also support the hypothesis of perception of partial cues. Under the conditions established to assess more carefully the presence of supraliminal cues, effects that might have been attributed to subliminal processes can be attributed to perception of supraliminal partial cues. In addition not only is there the generalization of the GSR to

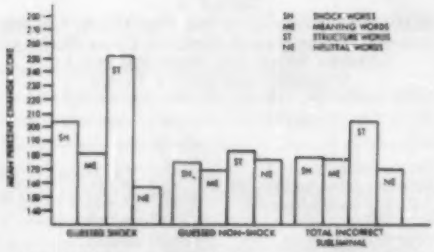

FIG. 2. Mean per cent change score of GSR for each stimulus category analyzed by type of responses for the incorrect subliminal trials.

structurally similar words (Fig. 2), but also the amount of response (Fig. 1) to the shock word decreased as the degree of subjective clarity of stimulus presentation decreased (i.e., from Supraliminal trials, to Correct Identification trials of Subliminal Presentation to Incorrect Identification trials of Subliminal Presentation). Further, the incorrect verbal "guessing" behavior indicated that the *S*s appeared to be responding verbally as well as autonomically to perception of portions of the stimuli.

The findings in this experiment bear upon another issue—whether or not subception in this paradigm is an experimental artifact (e.g., Ericksen 1956). The argument which suggests that subception effect may be an experimental artifact can be described as follows:

> Let us assume that conscious discrimination of the stimulus [verbal behavior] and the autonomic response evoked by the stimulus are correlated, but not completely, i.e., they do not correlate perfectly, even if both response systems were permitted to respond the same way, either as a continuous response or as a discrete response. [The issue is the same under conditions where one response is measured in discrete units, the other in continuous units.] This assumption requires only that it is possible for either or both responses to be right or wrong [by any criterion] or for one or the other to be right on any one trial. If this one assumption is tenable, then only four behavioral possibilities can result:
>
> (*a*) Verbal discrimination is correct—autonomic discrimination is correct
> (*b*) Verbal discrimination is correct—autonomic discrimination is incorrect
> (*c*) Verbal discrimination is incorrect—autonomic discrimination is incorrect
> (*d*) Verbal discrimination is incorrect—autonomic discrimination is correct

In research investigating the operations of subliminal processes, possibilities (*a*) and (*b*) are immediately excluded from further analysis because the organism is "aware" of (identifies) the stimulus and thus these trials are not considered as true subliminal presentation trials. Only possibilities (*c*) and (*d*) would be considered subliminal trials. Possibility (*c*) is not important for determining differences between the two response categories, because it contributes equal error variance to both systems—verbal and autonomic. Thus, since only (*d*) trials are contributing differences, given enough trials for (*d*) to occur, greater autonomic discrimination without verbal awareness could always be found. Because the conditions of analysis have been fixed, whether the stimuli are presented supraliminally or subliminally, autonomic responses will always seem to be more accurate than the verbal responses, and it will always be possible to conclude that subliminal processes perceive stimuli of which the organism is unaware, if the only trials analyzed include only trials on which the verbal (aware) response are incorrect. Whatever the merits of this analysis, the findings in this experiment suggest that the autonomic responses that occur and were interpreted as a subception effect need not be artifacts. Rather, they appear to be a generalization response to the perception of some part of the original conditioned stimulus. If they were artifacts, there would be no reason to expect the autonomic response to correlate more to one of the classes of stimuli (structure words) than to any of the other classes of stimuli (i.e., the other nonshock words, neutral or meaning) under the conditions of (*d*) trials.

EXPERIMENT III: SUBCEPTION OR PERCEPTION OF PARTIAL CUE IN THE PERCEPTION OF PERIPHERAL STIMULI PARADIGM

Some theorists who hold the two-process view maintain that the *S* "perceives" peripheral stimuli of which the *S* is apparently unaware at the time of exposure (e.g., Klein, 1956) but which have later effects upon the *S*'s behavior. Thus, for example, the concept of "registration" could be included in this discussion. That is, the stimulus affects the organism and is "received" by the *S*, but for some reason, does not reach the level of awareness, or consciousness, but still has some subsequent effect—the locus often unspecified. Klein (1956) describes the concept as follows:

> First the experiments show that *registration* must be distinguished from the different *experiential qualities* in which we may be aware of things and events. The processes responsible for phenomenal imprinting seem to be *independent* (this italics ours) of attention-giving processes—the means by which we become aware of things in a distinctive quality. Thus, registration and perception would seem to involve distinct processes (p. 11).

It would seem that registration for some theorists at least is considered a process different from, although an integral part of, perceptual processes in general.

On the other hand, if it is held, as a one-process theorist does, that effects attributed to phenomena such as registration (a paradigm of effects of unaware perception of peripheral stimuli) can be attributed to perception of partial cue, and are only quantitatively different from perception with awareness, it is necessary to demonstrate that effects attributed to "registration" can be shown to be related to the "amount" of partial cues perceived. This formulation of the problem essentially involves an incidental learning paradigm. Previous research has shown that the amount of incidental learning is affected by such variables as the type of orienting and peripheral tasks (e.g., Postman & Adams, 1956); the incentive of the *S* in performing the intentional task (e.g., Bahrick, 1954); and the conditions of presentation of the intentional material (e.g., Neimark & Salzman 1953). In this experiment the incidental learning (assumed to be a measure of the effects of perception) will be investigated under conditions of increasing difficulty of the intentional (i.e., orienting) task.

The third experiment, then, tests the hypothesis that the greater the amount of the peripheral stimulus *S* can perceive, the more registration occurs, even though the stimuli are objectively presented and perceptually equally available to the *S*s. In question in this hypothesis is the view that registration effects found are dependent upon supraliminal perception, as suggested by some theorists. Again, evidence that the effect of peripheral stimulation is a function of the amount of such stimuli perceived does not negate the possibility of registration as a phenomenon. Rather, it calls for closer consideration before the negative proposition is asserted, there is no evidence of supraliminal perception of the peripheral stimulation in registration. In this paradigm, the amount of peripheral stimuli available to the *S*s was varied by controlling the difficulty of an ongoing task, while the peripheral stimuli were exposed. The assumption was made that the more difficult a task (although not so demanding as to exclude any perception of peripheral cues), the less possible it would be to perceive peripheral stimuli.

Method

Subjects. Thirty undergraduate *S*s from an Introductory Psychology class were randomly assigned to one of three groups of 10 *S*s each.

Material. The test stimuli were nine irregular geometric forms. Two sets of these nine figures, randomly ordered (no design was repeated within each set of nine designs) were used. These 18 designs were pasted on a large circular board which rotated with the tracking boards. As the board rotated at a constant speed, each design appeared for a fixed time period in a window opening slightly above the place where the *S* was trying to track on the brass wire tracking pattern. Each of the three tracking boards had a different tracking pattern; one relatively straight and easy, the second more complex, and the third most complex and most difficult. The tracking board made eight complete constant speed revolutions in the 2 min. Each test figure appeared in the window opening 16 times for each *S*, and for the same amount of time for every *S*, no matter which task he was required to trace. The test sheet had the nine geometric designs which appeared on the rotor board interspersed randomly among 33 other irregular geometric patterns, all dittoed on a single sheet of 8½″ × 11″ paper.

Procedure. The *S* was seated in front of the pursuit rotor and the following instructions were read to him:

> I would like you to take this stylus. Behind these (pointing to the window covering) there is a window. In the window, you will see a brass line. When I start the apparatus, the brass line will start moving, and your task will be to keep the tip of the stylus on the brass line. The length of time you will be able to keep contact will be timed. [Pause] [Opening window cover] Do you see the brass line? O.K. Make contact with it now. Are you ready? Go.

After the *S* completed the pursuit rotor task, he was shown a sheet of paper on which were drawn 42 geometric figures, 9 of which were the same as the 9 exposed to the *S*s on the pursuit rotor task. The *S* was given the following instructions:

> While you were trying to keep the stylus on the brass line, did you notice that there were some figures going by? I'm going to give you a sheet of paper on which you will see a large number of figures. I would like you to pick out those nine figures which you think appeared up there.

TABLE 7
ANALYSIS OF VARIANCE OF THE NUMBER OF PERIPHERAL FIGURES CORRECTLY IDENTIFIED BY *Ss* IN THE THREE TASK GROUPS

Source	*DF*	*MS*	*F*
Between Conditions	2	19.00	6.418*
Within	27	2.96	

* $p < .01$.

Results

The score for each *S* was the number of correct designs selected from the 42 possible choices. An Analysis of Variance was computed for these scores for the three groups (i.e., Difficult, Moderate, Easy). Table 7 shows a statistically significant *F* value ($p < .01$) between the mean scores for the *Ss* in the various task groups. The mean correct identification for the Difficult Tracing Group was 1.8; Moderate 2.9; Easy 4.1.

Discussion

Again these results are consistent with the expectations derived from a one-process view and do not appear to be consistent with the expectations suggested by the two-process view that registration is a separate process. If the assumption is tenable that registration is independent of attention to the stimulus, as implied by some theorists, then the expectation that the amount of registration should be constant under equal stimulation seems to be a logical consequence. On the contrary, the results seem to show that tasks that demand greater attention and consequently less opportunity to perceive the peripheral stimuli results in less registration despite constant physical exposure, at least as measured by memory and identification of the peripheral stimuli. However as Klein has suggested, the effects may be more peripheral than the ability to identify the stimuli at a later time. Only additional research will determine whether these results will be found when other behavioral manifestations are explored.[7]

Conclusions

The findings of the three paradigm experiments are all consistent with a one-process view,—the hypothesis of "perception of partial cues." It is clearly recognized that these three experiments tap but a small range of the effects that have been attributed to subliminal processes. As Klein, Spence, & Holt, (1958) have noted, the effects of subliminal stimulation may be quite subtle and are not measured by the procedures used in these experiments. This latter issue is not in question. If subliminal effects occur they must be validly demonstrated by those who hold this view. What is relevant is that great care must be taken in accepting or asserting the negative proposition that no supraliminal effects are evident, involved in the inference of subliminal effects, without very carefully controlling and assessing the possibility that supraliminal effects *are* present. Greater care and subtlety must be used in ascertaining the presence or absence of supraliminal cues. Our results do *not* lead us to propose the view that there is *no* behavior without awareness, nor do we propose the hypothesis that only stimuli of which we are aware effect behavior. Rather, it is proposed that more careful and systematic exploration of behavior under varying levels of awareness be considered.

It is unfortunate that usage of the term "subliminal" and "awareness" has resulted in much confusion. The problem of subliminal has been discussed above. Consideration of the term "aware" shows equal difficulties and need for clarification. For example, one can be unaware of a stimulus (e.g., proprioceptive stimulation or X-ray); or be unaware of a response (e.g., a particular speech pattern); or unaware of the relationship between a stimulus and a response (e.g., conforming to norms of a normative group—aware of the behavior of the group, aware that we are making the same response, but unaware that conformity pressures have influenced our behavior). (Wiener, Carpenter, & Carpenter (1957); Wiener, 1959) Careful specification of these possible alternatives seems necessary. In addition, there is a tendency among some psychologists to postulate special and independent processes to account for behavior under low percentage threshold presentation conditions—and make these special instances of "unawareness," while finding no need to postulate special processes to account for the

[7] A further series of experiments are being prepared to test the effects of "registration" on less cognitive type behaviors, e.g., imagery, dreams, etc.

many other instances in which behavior without awareness has also apparently been demonstrated (e.g., Giddon & Ericksen 1959; Greenspoon 1955).

It would seem evident from these considerations and from the results of these experiments that investigations of "subliminal" and "unawareness" have not always carefully examined nor defined what was implied by these terms. To assume that a new or different process underlies variations in behavior under what appear to be variations in conditions, or to infer a new process based upon acceptance of a negative proposition may lead to a proliferation of constructs which will still further confuse and confound our limited understanding of behavior.

REFERENCES

ADAMS, J. K. Laboratory studies of behavior without awareness. *Psychol. Bull.*, 1957, **54**, 383–405.

BAHRICK, H. P. Incidental learning under two incentive conditions. *J. exp. Psychol.*, 1954, **47**, 170–172.

BLUM, G. S. Perceptual defense revisited. *J. abnorm. soc. Psychol.*, 1955, **56**, 24–29.

BRICKER, P. D., & CHAPANIS, A. Do incorrectly perceived tachistoscopic stimuli convey some information? *Psychol. Rev.*, 1953, **60**, 181–188.

BRUNER, J. On perceptual readiness. *Psychol. Rev.*, 1957, **64**, 123–152.

CARMICHAEL, L., HOGAN, H. P., & WALTER, A. A. An experimental study of the effect of language on the reproduction of visually perceived form. *J. exp. Psychol.*, 1932, **15**, 73–86.

ERICKSEN, C. W. Subception: Fact or artifact? *Psychol. Rev.*, 1956, **63**, 74–80.

GIDDON, N. S., & ERICKSEN, C. W. Generalization of response biases acquired with and without verbal awareness. *J. Pers.*, 1959, **27**, 104–115.

GOLDIAMOND, I. Indicators of perception: I. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychol. Bull.*, 1958, **55**, 373–411.

GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409–416.

JENKIN, N. Affective processes in perception. *Psychol. Bull.*, 1957, **54**, 100–127.

KLEIN, G. S. Consciousness in psychoanalytic theory: Some implications for current research in perception. Mimeo. paper read at Freud Centenary Symposium, Brandeis Univer., Massachusetts, 1956.

KLEIN, G. S., SPENCE, D. P., HOLT, R. R., & GOURE-VITCH, SUZANNE. Cognition without awareness: Subliminal influences upon conscious thought. *J. abnorm. soc. Psychol.*, 1958, **57**, 255–260.

LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness; A study of subception. *Psychol. Rev.*, 1951, **58**, 113–122.

MCCONNELL, J. V., CUTLER, R. L., & MCNEIL, E. S. Subliminal stimulation: An overview. *Amer. Psychologist*, 1958, **13**, 229–242.

MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.

MURDOCK, B. B., JR. Perceptual defense and threshold measurements. *J. Pers.*, 1954, **22**, 565–571.

NEIMARK, EDITH, & SALZMAN, I. J. Intentional and incidental learning with different rates of stimulus presentation. *Amer. J. Psychol.*, 1953, **66**, 618–621.

PERLOE, S. I. Inhibition as a Perceptual Process. Mimeo. paper read at APA, Cincinnati, 1959.

POSTMAN, L. J., & ADAMS, PAULINE A. Studies in incidental learning: IV. The interaction of orienting tasks and stimulus materials. *J. exp. Psychol.*, 1956, **51**, 329–333.

RIESS, B. F., Semantic conditioning involving the galvanic skin reflex. *J. exp. Psychol.*, 1940, **26**, 238–240.

RUBENFELD, S., LOWENFELD, J., & GUTHRIE, M. Stimulus generalization subception. *J. gen. Psychol.*, 1956, **54**, 117–182.

VOOR, J. H. Subliminal perception and subception. *J. Psychol.*, 1956, **41**, 437–458.

WHITE, C. T., & SCHLOSBERG, H. Degree of conditioning of the GSR as a function of the period of delay. *J. exp. Psychol.*, 1952, **43**, 357–362.

WICKENS, D. D. Conditioning to complex stimuli. *Amer. Psychologist*, 1959, **14**, 180–188.

WIENER, M. Perceptual thresholds: Conditions and parameters. Research proposal to United States Public Health Service, National Institute of Mental Health, 1957.

WIENER, M. Some correlates of conformity responses. *J. soc. Psychol.*, 1959, **49**, 215–221.

WIENER, M., CARPENTER, JANETH T., & CARPENTER, B. Some determinants of conformity behavior. *J. soc. Psychol.*, 1957, **45**, 289–297.

(Received August 3, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 138-140

CRITIQUE AND NOTES

TRUST, TRUSTWORTHINESS, AND THE F SCALE

MORTON DEUTSCH[1]

Bell Telephone Laboratories, Murray Hill, New Jersey

This study is part of a series of experimental studies (Deutsch, 1957, 1958) that have investigated some of the determinants of trusting behavior. In this brief paper, some striking findings relating trusting behavior to scores on the F scale (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950, p. 260) are reported.

Our experimental work has utilized a two-person non-zero-sum game (see Fig. 1) in which the gains or losses incurred by each person are a function of the choices made between two alternatives by one's partner as well as the choices made by oneself. Person I has to choose between Rows X and Y; Person II has to choose between Columns A and B. The amount of (imaginary) money each person wins or loses is determined by the box they get into as a result of their respective choices. For example, if Person I chooses Row X and Person II chooses Column A, they get into the AX box and they each win $9.

If you examine the possibilities of choice for Person I, you will notice that he can win most and lose least by choosing Y. Similarly, Person II can win most and lose least by choosing B. However, if Person I chooses Y and II chooses B, they both lose $9. Both can win only if they end up in the AX box. If Person I is reasonably sure that Person II is going to choose A, he can win more by choosing Y. Analogously, if Person II is confident that Person I is going to choose X, he can win more by choosing B rather than A.

The essential psychological feature of the game is that there is no possibility for "rational" individual behavior in it unless the conditions for mutual trust exist. If each player chooses to obtain either maximum gain or minimum loss for himself, each will lose. But it makes no sense to choose the other alternative, which could result in maximum loss, unless one can trust the other player.

In the study reported here, *S*s were drawn from an introductory psychology course at a local university. Several weeks prior to the experiment, they had filled out the F scale. During the experiment, the instructions to the *S*s about the game were such that they fully understood the implications of any combination of choices that they and the other person might make and they knew that the other person had a similar knowledge of the game. Unlike some of our other experiments (in which the *S*s were induced to assume a cooperative, individualistic, or competitive orientation), the *S*s were given no motivational orientation. They were allowed to assume whatever orientation they wished to assume vis-à-vis the other person, about whom they were given no information. The *S* did not know the identity of the other person and knew that the other person did not know his identity (except that each *S* knew that the participants in the experiment were all students in the same psychology course).

*S*s played the game twice, each time in a different "position" and each time, presumably, with a different person. In the First Position, *S* made his choice first and his choice was presumably announced to the other person before the other made his choice. In fact, the other person was "fictional" and, hence, the *S* was not informed what the "other person" chose after the *S* made his choice. In the First Position the *S* was faced with the decision of trusting the other person or not. In the Second Position, the *S* chose second after he knew the choice of the other person. Here, too, the other person was "fictitional" and the actual *S* was always informed that the "other person" had chosen Row X (i.e., had trusted). Hence, in the Second Position, the *S* was faced with the decision of being trustworthy or not.

Results

Table 1 presents the data concerning the relationship between choices in the two positions. It is clear that the *S*s who were "Trusting" when they chose first tended to be Trustworthy when they chose last; on the other hand, the *S*s who were "Suspicious" when they chose first tended to be Untrustworthy when they chose last.

Before the *S*s made their choices in the first position they were asked to indicate what they expected the other person to choose. Of the 24 *S*s who chose Row X in the first position and Column A in the second position, 21 indicated that they expected the other person in the second position to choose Column A; of the 22 persons who chose Row Y in

[1] This research was conducted while the author was on the faculty of New York University under a contract with the Office of Naval Research: Contract NONR 285(10). The United States Government is given permission to reprint this article in full or in part. I am grateful to Leonard Solomon for his assistance in collecting and analyzing the data and to Irving Taylor for making his students available for the experiment.

	A	B
X	(+9, +9)	(−10, +10)
Y	(+10, −10)	(−9, −9)

Person I chooses between Rows X and Y, Person II between Columns A and B. Person I's pay-offs are the first numbers in the parentheses, Person II's are the second numbers.

FIG. 1. Two-person non-zero-sum game.

the first position and Column B in the second position, 17 indicated that they expected the others in the second position to choose Column B. Thus, predominantly a choice of Row X was a "trusting" rather than "masochistic" choice and a choice of Row Y was correlated with an expectation of "untrustworthiness" rather than of "masochism" from the other person.

When the *S*s were in the second position they were asked to indicate what they thought the other persons (who were then in the first position) expected them to choose. The results indicate that 95% of the *S*s who were Trusting and Trustworthy in their own choices and who had also expected trustworthy choices from the other (i.e., a choice of Column A) thought the other person would be trusting (i.e., would expect a choice of Column A from the *S*); only 35% of the *S* who were Suspicious and Untrustworthy in their own choices and who had also expected untrustworthy choices from the others (i.e., a choice of Column B) thought the other person would be suspicious (i.e., would expect a choice of Column B from the *S*). This finding suggests that the *S*s who were Trusting and Trustworthy expected Trustworthiness and Trusting from the others while the *S*s who were Suspicious and Untrustworthy predominantly expected an exploitative orientation from the others (i.e., they expected the other persons to be trusting but not trustworthy). Hence, it is not surprising that, in the Second Position, these latter *S*s would respond to a Trusting choice from another by "taking advantage" of the other.

The data presented in Table 2 indicate a significant relationship between the *S*'s game behavior and his score on the F scale.[3] Almost all of the *S*s with relatively low scores made Trusting and Trustworthy choices, almost all of the *S*s with relatively high[3] scores made Suspicious and Untrustworthy choices, while the *S*s with medium scores tended to choose one way or the other equally often. It is interesting to note that 7 of the 9 *S*s who were inconsistent in their choices in the two positions (i.e., they were Trusting but Untrustworthy or Suspicious but Trustworthy) were in the medium range of the F scale scores for their class.

TABLE 1

RELATIONSHIP BETWEEN CHOICES IN THE FIRST AND SECOND POSITIONS

First Position Choice	Second Position Choice	
	Trustworthy (A)	Untrustworthy (B)
Trusting (X)	24	5
Suspicious (Y)	4	22

TABLE 2

RELATIONSHIP BETWEEN BEHAVIOR IN THE GAME SITUATION AND F SCALE

Game Behavior	F Scale Score			
	Low (1.3–2.2)	Medium (2.3–3.3)	High (3.4–4.4)	*N*
Trusting and Trustworthy	12	10	2	24
Suspicious and Untrustworthy	0	13	9	22
Suspicious but Trustworthy	0	4	0	4
Trusting but Untrustworthy	2	3	0	5

DISCUSSION

The data of Table 1 provide a striking demonstration of the symmetry of the *S*'s behavior in his two complementing roles vis-à-vis the other person. *His behavior toward the other is congruent with what he expects from the other, and also, what he expects from the other is congruent with his behavior toward the other.*[4] The *S*s tend to be trusting and trustworthy or suspicious and untrustworthy in this essentially ambiguous situation with unknown others. This result suggests that the personality predispositions tapped by the experimental game are not simply one-sided internalized orientations toward another or internalized expectations from another but are instead internalizations of a *reciprocal* pattern of interrelationships with another.

[3] The point biserial correlation between the scores on the F scale and the two most frequent categories of game behavior indicated in Table 2 is .50 ($p < .001$). The mean of the F scores for the Trusting and Trustworthy *S*s was 2.38; the mean for the Suspicious and Untrustworthy *S*s was 3.37.

[3] It is evident that the "highs" are high only in comparison with the other *S*s in their class; the high scores in this group are similar to the mean scores reported in other groups (Adorno, Frenkel-Brunswik, Levinson, & Sanford, p. 266). When using the F scale as a personality measure, we believe it makes most sense to consider the individual's score in relationship to the scores of others within his cultural milieu.

[4] It is apparent that the *S*s were not behaving in accordance with the ethical injunction of "do unto others as you would have others do unto you", but were rather guided by the dictum of congnitive consistency "do unto others as you expect others to do unto you and expect others to do unto you as you do unto them."

In other words, what appears to be internalized is a *system* of interrelations between oneself and the other, including the norms which prescribe *both* what to expect from the other and how to act toward the other. A similar conception of the nature of personality predispositions is advanced more fully by such authors as Mead (1934) and Parsons (1955).

The F scale data indicate that the game behavior was related to personality predispositions and that it was not determined by "accidental" orientations assumed during the course of the experiment, except in the special sense that vulnerability to "accidental" factors was itself influenced by the personality predispositions measured by the F scale.

The mass of research on the F scale, surveyed in Christie and Cook (1956), indicates that there are consistent differences between Low and High scorers in their behaviors in social situations and in their responses to various questionnaires. The High as compared with the Low scorers tend to be more authoritarian, less intellectually sophisticated, less liberal in their political views, more cynical concerning human nature, more prejudiced toward minority groups, and to have experienced and to favor stricter child rearing practices. The present results suggest that in addition, one may say that in an ambiguous situation involving the choices of trusting or not and of being trustworthy or not, Low scorers are more likely to be Trusting and Trustworthy while High scorers are more likely to be Suspicious and Untrustworthy (exploitatively oriented). These results are obviously concordant with the description of what the F scale was intended to measure and with the conception of the "antidemocratic personality" as advanced in *The Authoritarian Personality* (Adorno et al., 1950).

However, it is not necessary to posit as do the California authors (Adorno et al., 1950) that the characteristics of the authoritarian personality are to be

> Understood as expressions of a particular kind of personality structure within the personality. The most essential feature of this structure is . . . that the conscience or superego is incompletely integrated with the self or ego. . . . There is some reason to believe that a failure in superego internalization is due to weakness in the ego, to its inability to perform the necessary synthesis. . . . (p. 234)

An alternative viewpoint would be that the personality differences between High and Low scorers on the F scale do not necessarily reflect *structural* differences in personality so much as *content* differences in the values that have been internalized as a result of the individual's reaction to his socialization experiences in a particular social milieu, characterized by a given value pattern. This latter viewpoint suggests that Highs and Lows, Trusting people and Suspicious people do not necessary differ in superego integration or in ego weakness (a common synonym for "psychopathology"). As we have indicated more fully elsewhere (Deutsch, 1957), there are characteristic forms of pathological trust (e.g. "gullibility") as well as of pathological suspicion (e.g. "paranoia"). Presumably, the pathologies of trust and suspicion both reflect internal conflict and ego weakness, both of which may be found in individuals who have internalized widely differing values.

SUMMARY

Ss played an interpersonal game which, in one position, required them to choose between being trusting or suspicious of another and, in a second position, required them to choose between being trustworthy or untrustworthy toward another. There was a striking tendency for *Ss* who were trusting to be trustworthy and for *Ss* who were suspicious to be untrustworthy. F scale scores correlated significantly with game behavior; *Ss* with Low scores tended to be Trusting and Trustworthy while *Ss* with High scores tended to be Suspicious and Untrustworthy in their game choices.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harpers, 1950.

CHRISTIE, R., & COOK, P. A guide to published literature relating to the Authoritarian Personality through 1956. *J. Psychol.*, 1958, **45**, 171–99.

DEUTSCH, M. Conditions affecting cooperation. *USN tech. Rep.*, 1957, NONR-285(10). (Mimeo.)

DEUTSCH, M. Trust and suspicion, *J. conflict Resolut.*, 1958, **2**, 265–79.

MEAD, G. H. *Mind, self, and society,* Chicago: Univer. Chicago Press, 1934.

PARSONS, T., & BALES, R. F. *Family, socialisation, and interaction process.* Glencoe: Free Press, 1955.

(Received June 19, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 141–143

MOLAR VARIABILITY IN THE MENTALLY DEFECTIVE[1]

PAUL S. SIEGEL
University of Alabama

AND JAMES G. FOSHEE
Sunland Training Center

WE ARE concerned here with variability as it obtains among equally strong molar responses possessing instrumental equivalence. Upon successive presentations of a grossly constant stimulus occasion, variability among these alternatives may be observed to occur as an amalgamation of what Fiske and Rice (1955) have termed Type I and Type II variabilities. A description of the apparatus employed may provide clarification.

S is presented with the task of choosing among four switches. Approximately equal effort is required to execute each movement. All choices accomplish the common end of turning off a light, and it is assumed that *S* brings into the situation roughly equal response tendencies. A set is established through verbal instructions. *S* then responds to successive presentations of the light signal. Variability is expressed in the pattern of choices.

In this situation, a number of conditions may be varied. Within a factorial design, we have examined the effect of intellective impairment, sex, "motivation," and intertrial delay. But, for present purposes, our concern is with intellective impairment. How does the behavior of the mentally retarded differ from that of the "normal"?

METHOD

Subjects and Apparatus

The experiment employed 128 *Ss*. Sixty-four were recruited from the Cloverbottom Home, an institution for the mentally defective. Both familial and organic types were represented. This distinction was felt to be unnecessary for the purpose of the study. Gross sensory and motor disturbances were excepted. Chronological ages ranged from 11 to 46 yr. with a median of 28. Half the *Ss* were male. Mental ages (Stanford-Binet) ranged from 4–0 to 10–2 yr. with a mean of 6–7 and a *SD* of 15.1 mo. Seventeen of the 64 *Ss* were on a mild drug regimen (dilantin, serpasil, or phenobarbital). They were randomly assigned to conditions. To establish the absence of a confounding drug effect, the performance of these *Ss* was separately analyzed upon the conclusion of the experiment. Two comparisons were made. The mean variability score (39.6) of this subgroup was compared with the mean of the total group (37.9) and with the mean (36.5) of 17 nondrugged mental defectives chosen at random from the remaining *Ss* within each matching condition. In both instances, the difference is readily attributed to chance. It is felt, therefore, that these *Ss* were representative in their task performance, and their results have been included in the final analysis.

The other 64 *Ss* were drawn from the Peabody Demonstration School and the Tennessee Orphan Home. The Lorge-Thorndike Intelligence Test (nonverbal battery) was utilized to establish normal intelligence. Half the *Ss* were male. Chronological ages ranged from 5–10 to 9–8 with a median of 8–4. Mental age equivalents ranged from 5–0 to 12–2 yr. with a mean of 8–8 and *SD* of 23 mo. The designation of a control group for an institutionalized, mentally defective population is necessarily done at the cost of introducing some element of arbitrariness. We had hoped to realize a closer match for mental age, but difficulties arising in the recruitment of a "normal" group dictated otherwise. It should be noted that our "normals" possessed a higher mean mental age and manifested greater dispersion in this characteristic.

In detail, the apparatus consisted of a half-inch plywood board, painted red, upon which were mounted four micro-switches (hereafter, arbitrarily designated from left to right as 1, 2, 3, or 4) and a light (7½ w.). The board was tear drop in shape with a long dimension of 28 in. The diameter of the rounded end was 21 in. The switches were arranged in a semicircle about a central black spot, 5 in. in diameter. *S* was seated in a chair before a small table. The board was then positioned by *S* to provide equal access to all four switches with the forearm resting upon the narrow end of the board. Each switch was of yellow plastic, was mounted atop the board, and possessed ¼-in. shaft which projected about ¼-in. above the switch body. Depression of the shaft generated a clearly audible click and served to extinguish the light. The light bulb was mounted behind a translucent pane, 1 in. in diameter, within a small wooden compartment

[1] This study was undertaken while the senior author was serving as Visiting Professor at Vanderbilt University in the summer of 1957. In part it was supported by the Research Committee of the University of Alabama. The authors wish to acknowledge the assistance of Mr. Ben Allen in gathering the data.

[2] Fort Myers, Florida.

TABLE 1
MEAN INDEX OF VARIABILITY OBTAINED UNDER VARIOUS CONDITIONS

	High Motivation				Low Motivation			
	(Short Delay)		(Long Delay)		(Short Delay)		(Long Delay)	
	M	*F*	*M*	*F*	*M*	*F*	*M*	*F*
Mentally Defective	37.3	41.5	34.5	39.3	36.0	43.5	35.0	36.0
Normal	28.4	33.5	35.7	31.0	22.5	37.0	20.0	28.3

which extended beyond the tear drop in a position midway between switches two and three.

As one condition, threat of shock was utilized as a "motivater." A zinc electrode was strapped to the nonpreferred hand and the lead was plugged into a nonfunctional GSR apparatus located within *S*'s field of vision.

Instructions to *S* were as follows:

> Please stand here. When this light comes on (*E* turns on the light), I want you to turn it off. Any one of these buttons will turn off the light. Watch me. See, I always start from here (*E* places his hand on the black spot) and when I turn the light off, I put my hand on this spot (*E* operates the switches in a 4-3-2-1 order). Now you try it (*S* is then seated). Move the chair until you are comfortable and can reach all the buttons easily. Let's do a few for practice. Try this one; now this one; now this one; and now this one (*E* directs *S* to use the switches in a 2-4-3-1 order and insures that *S* is centered).
>
> O.K. Do you understand what I want you to do? Remember, when the light comes on, you turn it off. Use *any* button you want to. Always bring your hand back to the spot after pushing a button. Ready?"

When threat of shock was employed, the following instructions were added:

> Turn it off as fast as you can. If you wait too long to turn the light off, you may get a little electric shock. The shock won't hurt but it will give you a funny feeling. Just hurry each time and you will *not* get shocked at all. Ready?

Each *S* was then given 33 successive trials; a trial consisting of light presentation followed by response. Switch choice was recorded by *E*, seated to the left and slightly behind *S*.

The major subgroups, mentally defective and normal, were equally split into male and female subgroups. In turn, these subgroups were split equally into high (threat of shock) and low (no threat) motivation subgroups. Short (2 sec.) and long (30 sec.) intertrial delays provided the basis for a fourth split in the design. Table 1 depicts the design and presents the outcome.

RESULTS

The 33 responses yielded by each *S* were tabulated in a 4 × 4 contingency table. Each successive pair of responses defined an entry. If, for example, on the first four trials, *S* yielded the successive choices, 3-3-1-4, an entry was made in the third column and third row cell, the third column and first row cell, and the first column and fourth row cell. *S*s persisting in the choice of a single switch yielded 32 entries of a single cell. *S*s staying with a two-switch pattern—say, 1-2-1-2-1-2, etc.—yielded 16 entries each in two cells. *S*s consistently following a four-switch pattern—say, 1-2-4-3-1-2-4-3, etc.—yielded eight entries each in four cells. The upper limit of randomness, variability, or response independence would be reflected in two entries each of 16 cells.[3]

From this contingency table, a variability index was computed for each *S*. This index was arrived at by summing the absolute deviations from the expected random value of two for each cell. The deviations were not squared (chi square) because this operation produced a marked skewing of the distributions. The simple sum was taken as an index of variability or response dependence with low scores indicative of high variability or response independence and high scores of low variability or response dependency. The lowest obtained index was 10 (0 was possible), the highest, 60 (yielded by *S*s persisting in the choice of a single switch). This index was found to be distributed in a reasonably normal fashion, and conventional variance analysis was applied.

The mean value obtained from the eight *S*s examined under each set of conditions is reported in Table 1. As a total group, the mentally defective yielded a mean index of 37.9, the normal 29.6. An *F* of 21.34 indicates a high degree of reliability (*P* is far less than .01). We conclude that in the present experimental task, the mentally defective express significantly less variability (or greater response dependency). For the purpose of this study, we may neglect other primary effects. Interaction of intellective impairment with other "dimensions" is of concern, but no interactions were found to be statistically significant.

DISCUSSION

Lewin's early analysis of the feebleminded (1935) permits a kind of intuitive grasp of the present findings. Response variability possesses face validity as an index of "rigidity." This position, like the related one of Kounin (1941a, 1941b), predicts that the more rigid mental defective will show less variability in the present task.

[3] Emil W. Menzel pointed out the applicability of this analysis to the present data.

Even a cursory reading of the literature of the past few years reveals that the concept of rigidity has fared badly. As an alternative, the present investigators wish to propose an S-R account. The principle of disinhibition presents attractive possibilities. Admittedly invoked *ad hoc*, this mechanism can "explain" the outcome of this study. It may possess broader scope as a theoretical account of the behavior of the mental defective. Since its implications are testable, its generality can be experimentally assayed.

If we assume that some sort of inhibition is laid down with the execution of a two-choice pattern—say, 1 followed by 3—any condition that serves to disrupt or dissipate this state will facilitate the repetition of that pattern. Variability is decreased, and for a series of choices, the overall index of response dependency is increased. The mental defective, manifesting a kind of "scatter" to his behavior, is commonly described as "distractible." This invites the present interpretation that his lowered variability results from a more pronounced disruption of inhibition, arising out of his lessened capacity to shut out even minor distraction. This view is analogous to that offered by Hovland (1936) to account for an "anomalous" upsurge of the extinction curve on the initial trial following the cessation of reinforcement. Hovland felt that the sudden shift from reinforcement to nonreinforcement constituted a condition sufficient to dispel the "inhibition of reinforcement" accumulated during massed learning trials. We see the mental defective as peculiarly susceptible to such disinhibiting influences, hence given to lowered variability or repetitiousness when held within a task framework.

The inhibitory state assumed here can be variously identified as reactive inhibition (Hull, 1943), as stimulus satiation (Glanzer, 1953), or as central in the sense used by Duncan (1956). It makes little difference. We assume only that responding, perceiving, or, broadly, experiencing generates a condition at least momentarily inhibitory to the repetition of the experience. We assume further that this inhibitory state is susceptible to external influence and that greater vulnerability is a distinctive characteristic of the mentally defective. In the present study, the normal flux of the environment was taken as disinhibiting. As a deduction from our interpretation, the effect of controlled stimulation upon the variability of the mentally defective might be examined.

It would be desirable to generalize the present findings to other tasks, to other controls, and to other indices of variability. The task was a loosely governed one. Selection of the control group was determined largely by convenience. And the index of variability possesses elements of arbitrariness. The present study must be viewed as exploratory and as hypothesis-generating, not as hypothesis-testing.

SUMMARY

Mental defectives and normal children were required to choose among four switches. All choices accomplished the common end of extinguishing a light, and approximately equal effort was involved in the execution of each choice or response. In a series of successive responses, the mental defective manifested greater response dependency or less variability.

This finding was placed within the framework of disinhibition theory. It was reasoned that the lowered variability of the mental defective results from a more pronounced disruption of the inhibition to repeat arising out of a lessened capacity to exclude minor distractions.

REFERENCES

DUNCAN, C. P. On the similarity between reactive inhibition and neural stimulation. *Amer. J. Psychol.*, 1956, **69**, 227–235.

FISKE, D.W., & RICE, LAURA. Intra-individual response variability. *Psychol. Bull.*, 1955, **52**, 217–250.

GLANZER, M. Stimulus satiation: An explanation of spontaneous alternation and related phenomena. *Psychol. Rev.*, 1953, **60**, 257–268.

HOVLAND, C. I. "Inhibition of reinforcement" and phenomena of experimental extinction. *Proc. Nat. Acad. Sci., Wash.*, 1936, **22**, 430–433.

HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.

KOUNIN, J. S. Experimental studies of rigidity: I. The measure of rigidity in normal and feeble-minded persons. *Charact. Pers.*, 1941, **9**, 251–272. (a)

KOUNIN, J. S. Experimental studies of rigidity: II. The explanatory power of the concept of rigidity as applied to feeble-mindedness. *Charact. Pers.*, 1941, **9**, 273–282. (b)

LEWIN, K. *Dynamic theory of personality*. New York: McGraw-Hill, 1935.

(Received May 14, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 144-145

IDENTIFYING JEWISH NAMES:

DOES PREJUDICE INCREASE ACCURACY?[1]

PAUL F. SECORD AND ELLEN SAUMER
University of Nevada

AN EARLY study by Allport and Kramer (1946) indicated that anti-Semitic persons were able to identify photographs of Jews and non-Jews with a higher degree of accuracy than less prejudiced persons, and that the high prejudice group had a tendency to label more photographs as Jewish than the low group. Allport and Kramer (1946) theorized that those who are prejudiced against a minority group are sensitized to the visible signs of identity of members of such groups. Although their study was supported by Lindzey and Rogolsky (1950), Carter (1948) found no relationship between degree of prejudice and accuracy of identification. Moreover, Elliott and Wittenberg (1955) subsequently presented evidence that the relationship found in former studies was an artifact due to response bias, i.e., the tendency of prejudiced individuals to label photographs as Jewish. Scodel and Austrin (1957), in a study designed to eliminate the effects of response bias upon accuracy, found no relationship between anti-Semitism and accuracy of identification of Jews. Like others before them, they also demonstrated that prejudiced individuals do indeed tend to identify more photographs as Jewish.

These studies employed Jewish and non-Jewish photographs to provide cues to identification of minority group membership. Physiognomy, however, provides only a limited variety of cues to identity. If persons with anti-Semitic attitudes become sensitized to the signs of the identity of Jews, they should also be able to utilize cues other than physiognomic ones. An ever-present and familiar cue to minority group identification is a person's name. The purpose of the present study is to determine whether there are differences between high and low prejudiced groups of gentiles on response bias and accuracy of identification of Jewish and gentile names.

METHOD

A Jewish name was operationally defined as one which is held by a person who states, explicitly or implicitly, that he is a Jew. Fifty Jewish names were taken at random from the roster of approximately 600 delegates to the 1953 American Jewish Conference. Twenty-five additional names were taken at random from a directory and were designated as non-Jewish names. None of these names was the same as any name in the sample of Jewish names. The names were combined into one list[2] which was presented to two groups: college students and United States Army personnel. The former consisted of 67 students obtained from three lower division psychology classes at the University of Nevada, and the latter of 197 soldiers at Fort Ord, California.

Since Elliott and Wittenberg (1955) discovered that under certain circumstances accuracy scores are a function of response bias, two procedures were employed. One of these measured response bias, and the other, accuracy uninfluenced by response bias. *Ss* were first asked to designate each name on the list as either Jewish or non-Jewish. The frequency with which they designated names as Jewish represented their response bias. When this procedure had been completed they were given another copy of the list and asked to designate exactly 50 of the 75 names as Jewish. The frequency with which they chose actual Jewish names constituted their accuracy score.

Finally, in order to obtain an index of prejudice, *Ss* were given a questionnaire consisting of 26 items selected from the California A-S Scale (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950), a measure of anti-Semitism. After completing the questionnaire, *Ss* were asked to indicate whether they were gentile or Jewish. Data obtained from those who designated themselves as Jewish, or who failed to follow instructions on any of the forms were discarded. The 65 *Ss* discarded for the Army group, a large number, did not vary in any systematic way from the retained sample. The greatest single category of error was in failing to designate exactly 50 names in the accuracy variation—an error occurring in spite of carefully worded instructions requiring *Ss* to make two counts. *Ss* who checked 49 or 51 names were retained, how-

[1] The authors wish to thank Howard H. McFann, who made arrangements for obtaining United States Army personnel as subjects, and Robert D. McDonald, who administered the forms to the military group.

This investigation was supported in part by a grant (M-1892) from the National Institute of Mental Health, Public Health Service.

[2] The 75 names have been deposited as Table A with the American Documentation Institute. Order Document No. 6236 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 1
ACCURACY AND RESPONSE BIAS AS A FUNCTION OF PREJUDICE

Prejudice	Accuracy		Response Bias	
	Mean	Sigma	Mean	Sigma
College students				
High (N = 31)	38.61	2.4	26.03	8.9
Low (N = 31)	38.29	2.7	26.16	8.9
Mean difference	+0.32*		−0.13*	
Army personnel				
High (N = 66)	36.00	2.7	33.02	10.0
Low (N = 66)	36.56	2.6	27.85	12.0
Mean difference	−0.56*		+5.17*	

* $P > .05$ level.

ever, with E adding or deleting a name by a random method. Only five college students were deleted, and three of these were discarded because they were Jewish.

RESULTS

"High" and "low" prejudice groups were formed by designating those above the median A-S score for that sample as "high" and the remainder as "low." Since only nine of the 62 A-S scores for college students were in a prejudiced direction from the "neutral point" of the scale, and slightly more than half of the Army personnel had scores in the prejudiced direction, results were treated separately for the two samples. High and low prejudice groups for each sample were compared on accuracy and response bias as reported in Table 1. No significant differences ($P > .05$) between the high and low prejudice groups were found on either response bias or accuracy. As an additional check, means were also computed for the highest quartile and the lowest quartile. Results were essentially the same as those reported in Table 1, with no appreciable or significant differences occurring.

When names were ranked according to the frequency with which they were designated as being Jewish, it was apparent that many were overwhelmingly chosen as Jewish and many rarely so designated. Since, for statistical reasons, names always or rarely identified as Jewish cannot discriminate as readily between high and low prejudice groups, the high and low groups were compared for accuracy on the 25 names in the middle frequencies. Again, however, there was no significant difference between the two college groups (mean diff. = −0.12), or the Army groups (mean diff. = +0.18).

Finally, the question was raised as to whether the obtained mean frequency of approximately 38 correctly identified Jewish names (36 for Army personnel) was significantly above the theoretical frequency of approximately 33 names which could be obtained by chance. A one-tailed binomial test (Siegel, 1956) comparing these frequencies, indicated that they were not significantly different (college Ss, $P > .11$; Army Ss, $P > .26$).

DISCUSSION

Data obtained here are consistent in every way with the conclusion that there is no difference between high and low prejudice groups either in their ability to discriminate Jewish from gentile names or in their tendency to label names as Jewish. Response bias has usually been obtained in previous studies with Jewish photographs; the present results differ largely in failing to obtain response bias for Jewish names. The reason for the failure to obtain a significant degree of response bias is not clear.

The present negative results for accuracy of identification of names are consistent with the fact that properly controlled photographic studies have been uniformly negative. It seems reasonable to conclude that the hypothesis that prejudice increases accuracy of identification of Jews may be safely put to rest, with only a remote possibility that an exhumation will later be called for.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

ALLPORT, G. W., & KRAMER, B. M. Some roots of prejudice. *J. Psychol.*, 1946, **22**, 9–39.

CARTER, L. F. The identification of "racial" membership. *J. abnorm. soc. Psychol.*, 1948, **43**, 279–286.

ELLIOTT, D. N., & WITTENBERG, B. H. Accuracy of identification of Jewish and non-Jewish photographs. *J. aborm. soc. Psychol.*, 1955, **51**, 339–341.

LINDZEY, G., & ROGOLSKY, S. Prejudice and identification of minority group membership. *J. abnorm. soc. Psychol.*, 1950, **45**, 37–53.

SCODEL, A., & AUSTRIN, H. The perception of Jewish photographs by non-Jews and Jews. *J. abnorm. soc. Psychol.*, 1957, **54**, 278–280.

SIEGEL, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.

(Received March 30, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 146-147

TAYLOR MAS AND GROUP CONFORMITY PRESSURE

G. L. MANGAN, D. QUARTERMAIN, AND G. M. VAUGHAN
Victoria University, Wellington, New Zealand

EXPERIMENTAL studies of conformity in a situation of group pressure have increased rapidly in number since Asch's (1952) original demonstration. Such studies suggest that the determinants of conforming behavior might appropriately be classified under three headings involving (*a*) the task stimulus (ambiguity, for example), (*b*) the social influence (number of confederates), (*c*) the characteristics of the individual (personality correlates).

In a previous study (Mangan, Quartermain, & Vaughan, 1959) we investigated a variable of the third type, namely the relationship between conformity scores in a setting like Asch's and scores on the Taylor Scale of Manifest Anxiety (Taylor, 1953). The results showed no significant difference in gross amount of yielding between *Ss* who scored high (HA) and those who scored low (LA) on the Taylor Scale. However, a secondary analysis of yielding responses in terms of the *Ss*' position in the response sequence, which was taken as a measure of the amount of pressure being exerted on the *S*, showed that the HA *Ss* conformed less, and the LA *Ss* more, as they responded later in the sequence. We suggested that, as pressure increases, the HA *Ss* show an increasing amount of hostility and a resulting unwillingness to conform.

In the present research, we attempt to investigate this effect more directly. We hypothesized that when group pressure is severe, *Ss* low in Taylor anxiety should give significantly more yielding responses than *Ss* high in Taylor anxiety.

METHOD

Material and Subjects

The experimental material consisted of 20 slides, measuring 2¼ in. x 2¼ in., each mounting two clusters of dots, labelled A and B, which were clearly differentiable.

The 12 highest and the 12 lowest scorers on the shortened 28-item Taylor MAS (Taylor, 1953) were selected from a total of 160 undergraduate students as *Ss* for this experiment. The mean Taylor score for the LA *Ss* was 3.25, for the HA *Ss* 18.25. The confederates were 12 specially selected undergraduate students, who underwent a rigorous training procedure before the experiment was begun.

Procedure

The four confederates and the *S* were instructed to be outside the experimental room at a certain time. The confederates approached from different directions and involved the *S* in a discussion about where the experiment in psychology was to take place and when. The experimenter then appeared and ushered the group into the experimental room. The confederates arranged for the *S* to sit at one end of the semicircle of chairs, which were placed immediately in front of the projector and about five yd. from the screen.

The group was given the following instructions:

> This is an experiment in perceptual speed. I am going to show you some slides for a short exposure. On these slides you will see two groups of dots, labelled A and B. I want you to tell me which of the two groups contains the larger number of dots.

The experimenter exposed each slide for 1 sec., and responses were recorded. The order of response was rotated in each trial so that each *S* responded in each of the five positions on four occasions. The predetermined responses of the confederates were coded and embedded in a large block of data written upon the blackboard in the experimental room under the legend "Stage III results, Experiment 15: April 10th." "Please leave" was written across the table.

Four of the 20 trials were considered noncrucial—i.e. in these the confederates gave the correct response without comment. In ten of the remaining 16, the confederates gave incorrect responses without comment. In the remaining six trials, in which the *S* responded third (Trials 8 and 18), fourth (Trials 9 and 19), and fifth (Trials 10 and 20) in the response order, the following procedure was adopted. The first confederate on these six occasions answered correctly in a hesitant manner. He was immediately criticized strongly by the confederate or confederates following him, but preceding the *S* in the response sequence. These gave the incorrect response. The first confederate appeared confused and then conformed. In these trials, therefore, pressure was exerted indirectly on the *S* before he was required to respond. It was assumed that the criticism levelled at the first confederate and his subsequent conforming behavior constituted pressure on the *S* about to respond. To avert suspicion about the object of the experiment it was considered essential to limit the number of these "strong pressure" trials to six.

At the end of the session the *E* disclosed the true nature of the experiment to the *S*.

RESULTS

The 12 HA *Ss* gave 36 yielding responses (19%) compared with 64 (33%) for the 12 LA *Ss*. This difference in gross amount of yielding, when tested by a Mann-Whitney *U* Test, is significant at better than the .001 level.

The 16 crucial trials were ordered into three categories: the six strong pressure trials, in which there was interchange of comment between the confederates, the three pressure trials immediately

TABLE 1
PERCENT OF SUBJECTS YIELDING IN THE THREE CATEGORIES OF TRIALS

Category of Trial	Percentage Yielding	
	LA	HA
Strong Pressure Trials	25	8
Immediately following Trials	50	28
Pressure Trials	35	23

following the strong pressure trials and the remaining seven pressure trials in which the confederates gave the incorrect responses without comment. Percentages of yielding responses in the three blocks of trials are presented in Table 1.

The differences in yielding percentages in the strong pressure and the immediately following trials are significant for both the LA *Ss* ($P < .006$) and the HA *Ss* ($P < .04$) by the sign test. None of the other differences is significant, although a comparison of yielding percentages of HA *Ss* in the strong pressure and the pressure trials is suggestive ($P < .07$).

DISCUSSION

Judging from Asch's original data (32%) and from those reported in the previous study (33% for the LA *Ss*, 27% for the HA *Ss*) it is clear that the difference in yielding between the HA and LA *Ss* in this experiment is due to the low conformity of the HA *Ss*. The LA *Ss* conform with about the same frequency as previously reported. The original prediction therefore is strongly supported.

A clear trend is evidenced in Table 1. Both groups, particularly the HA *Ss*, show very much lower yielding percentages for the six strong pressure trials, and both groups show significantly more receptivity to pressure in the three immediately following trials, the latter effect being stronger among the LA *Ss*.

The reactions of the HA and the LA *Ss* to the experimental procedure were strikingly different. The LA *Ss* appeared disoriented and confused and often made self-disparaging remarks and asked for instructions, particularly in the strong pressure trials. On the other hand, the HA *Ss* seemed somewhat suspicious, occasionally criticized the confederates and showed signs of anger or hostility or both. This was clearly discernible in the post-experimental enquiries. We include a brief description of one such enquiry each, from an HA and an LA *S*.

Q. Did you think there was anything strange about the whole setup?

LA	HA
A. No. I thought after a while that I'd made a mistake and that I was supposed to name the group having the smaller rather than the larger number of dots. [At this point, many LA *Ss* asked for instruction.] After that, I was rather confused.	A. Probably; weren't the others playing follow the leader—trying to have me on or something?

Q. Did you think there was any collusion between the *Ss*?

A. At one stage I had a vague suspicion, but didn't really know.	A. Generally, yes. At times they sounded very convincing and I wasn't at all sure; why should I change my mind anyway no matter how many think differently.

Q. Did you think there was any collusion between the *E* and the other *Ss*?

A. No.	A. No.

The impression gained from these enquiries and from observations made during the course of the sessions is that whereas the LA *Ss* tend to question their own judgment more but still to show "normal conformity," under strong pressure conditions the HA *Ss* become somewhat hostile and rigid and less amenable to conformity influence. Some general support is thus given to the findings of Janis and Feshbach (1953) that increase in emotional tension reduces the effectiveness of persuasive communications. More specifically, these results substantiate Janis' (1954) original hypothesis that *Ss* who show marked neurotic anxiety symptoms as measured on selected items of the MMPI and the United States Army Neuropsychiatric Screening Adjunct are less compliant to persuasion than those *Ss* who score lower on these items. It must be pointed out, however, that Janis failed to confirm his earlier hypothesis in a later study (Janis 1955) in which he used a questionnaire derived by Sarason & Gordon (1953) as a measure of neurotic anxiety.

REFERENCES

ASCH, S. E. *Social psychology*. New York: Prentice Hall, 1952.

JANIS, I. L. Personality correlates of susceptibility to persuasion. *J. Pers.*, 1954, **22**, 504–518.

JANIS, I. L. Anxiety indices related to susceptibility to persuasion. *J. abnorm. soc. Psychol.*, 1955, **51**, 663–667.

JANIS, I. L., & FESHBACH, S. The effects of fear arousing communication. *J. abnorm. soc. Psychol.*, 1953, **48**, 78–92.

MANGAN, G. L., QUARTERMAIN, D., & VAUGHAN, G. Relationship between Taylor MAS scores and group conformity. *Percept. mot. Skills*, 1959, **9**, 207–209.

SARASON, S. B., & GORDON, E. M. The Test Anxiety Questionnaire scoring norms. *J. abnorm. soc. psychol.*, 1953, **48**, 447–448.

TAYLOR, J. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

(Received May 21, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 148-150

PERSONALITY AND NEGRO-WHITE INTELLIGENCE[1]

SHELDON R. ROEN
Unive sity of New Hampshire

THE SUPREME COURT in its decision on school integration found that segregation in the public schools implies the inferior status of Negro children thereby retarding their "educational and mental development." This ruling has evoked considerable controversy among psychologists. McGurk (1956), in a widely circulated magazine article purporting to survey "the *only* existing studies that relate to the problem" (p. 96) so as to make "sense" out of the Supreme Court's decision, concluded that "there is something more important, more basic to the race problem than differences in external opportunity" (p. 96). He implied that there was a contradiction between the Supreme Court's decision and the psychological literature. Shuey (1958) in a book reviewing the research on Negro intelligence concludes that the evidence "points to the presence of some native differences between Negroes and whites as determined by intelligence tests" (p. 318). Reactions to this resurgence of racist theory (Klineberg, 1957; Long, 1957; North, 1956) stress the lack of justification for generalizing about race differences, but add no new data.

It seems that a rather important dimension, that of personality as it is manifested in self-perceptions, has been overlooked. A search of the literature reveals that there is evidence: (*a*) that the personalities of Negroes are different from whites (Boyd, 1952; Karon, 1957); (*b*) that personality is related to ability to perform on intelligence tests (Sontag, Baker, & Nelson, 1955; Trumbull, 1953); (*c*) that as infants (Gilliland, 1951) and preschool children (Anastasi & D'Angelo, 1952), Negroes do not score lower on measures of intelligence (these two studies were overlooked by Shuey and by McGurk); and (*d*) that the emergence of personality is closely associated with environment. However, no studies were found that relate the points to one another.

Viewing the issue in this way suggests the proposition that indirect psychological experiences of socioeconomic exclusion, complicated by feelings of lack of historical achievement, and generally more erratic family ties, negatively influences the emerging personality or self-perceptions of Negroes, especially as these relate to their intellectual potentials. Before the proposition can be dealt with directly, the relationship of certain personality traits to intellectual variables in Negroes should be established. The following hypotheses were formulated: (*a*) Negroes would score lower on measures of intelligence than would whites; (*b*) Negroes would score less positively on measures of certain personality traits than would whites; and (*c*) personality variables would correlate more highly with intelligence test scores in Negroes than they would in whites.

METHOD

The personnel files of hundreds of soldiers stationed in an Army post in the Southwest were examined for biographical data and Army Classification Battery (ACB) scores. About 500 were selected for testing and interviews by eliminating those of extremely low ACB scores and those whose biographical data would be hard to duplicate. Personality tests available for administration were the California Test of Personality (Thorpe, Clark, & Tiegs, 1953), the Taylor Manifest Anxiety scale (Taylor, 1953), and the Bernreuter Personality Inventory (Bernreuter, 1935). A plea was made for the soldiers' cooperation and those not desiring to participate were excused. The tests were taken under conditions of anonymity and 98% cooperated.

A measure of test-taking "integrity" was developed by selecting items on the tests that on face value (judgments of three clinicians) appeared to be contradictory. *S*s who responded in a contradictory manner to three or more pairs of items were eliminated from the sample.

Fifty Negro and 50 white soldiers were then closely matched on 10 variables: age, education, occupation of parents, income of parents, geographical area of childhood, army rank, number of years in the service, marital status, urban or rural background, and plans for re-enlistment. Matching was accomplished by categorizing and subcategorizing the variables and then pairing a Negro with a white soldier who most closely resembled him in biographical pattern. A typical example of a matched pair was where both: (*a*) were married and between 25 and 35 years old, (*b*) graduated from high school but did not enter college, (*c*) had fathers who were unskilled laborers, (*d*) indicated that the income of their parents was below $5,000 a year, (*e*) were brought up in the Northeast

[1] A modification of this paper was read at the APA convention in Cincinnati, 1959.

The analysis of the data for this research was supported by the Central University Research Fund of the Graduate School at the University of New Hampshire.

TABLE 1
DIFFERENCE BETWEEN MEAN SCORES FOR NEGROES AND WHITES

Measure	Negro		White		p
	M	*SD*	*M*	*SD*	
Combined intelligence	91.3	16.5	101.2	19.0	<.005
Pattern analysis	89.2	20.2	103.0	21.6	<.005
Total adjustment	127.7	17.5	129.2	20.6	*NS*
Manifest anxiety	13.8	11.2	12.9	8.3	*NS*
Lack of self-confidence	5.6	83.6	−43.0	67.6	<.005

TABLE 2
DIFFERENCES BETWEEN CORRELATIONS FOR NEGROES AND WHITES

Personality Measure	Combined Intelligence			Pattern Analysis		
	Negro	White	p	Negro	White	p
Total Psychological adjustment	.45**	.35**	*NS*	.73**	.22*	<.001
Manifest anxiety	−.55**	−.27*	<.05	−.40**	−.23*	NS
Lack of self-confidence	−.79**	−.50**	<.01	−.88**	−.46**	<.001

* Significant beyond the .05 level.
** Significant beyond the .01 level.

section of the United States and in cities over 500,000 in population, (*f*) were sergeants, (*g*) were in the army from 6 to 10 years and planned to reenlist. Although this type of matching might tend to mask some important personality differences that might be found in representative Negro and white groups, the desirability of careful matching for the purpose of obtaining comparable correlations between measures of intelligence and personality in the two groups seemed to outweigh the likelihood of obscuring such group differences in personality.

Two measures of intelligence were derived from ACB scores. The first measure, the "combined intelligence measure", was obtained by averaging the scores on three tests that form part of the Army Classification Battery. These tests were found (Montague, Williams, Lubin, & Gieseking, 1957) to correlate with the Wechsler-Bellevue as follows: RV (reading and vocabulary) .77, AR (arithmetic reasoning) .78, and PA (pattern analysis) .81. The PA score, since it correlates highest with the Wechsler-Bellevue, was used as a second measure of intelligence.

RESULTS

On both the combined intelligence measure and the Pattern Analysis test, the Negro group scored significantly lower than the white group. The mean Negro score was 91.3 for the combined measure, and 89.2 for pattern analysis. The mean white score was 101.2 for the combined measure and 103.0 for pattern analysis.

Three scores were used to measure personality variables: (*a*) the total adjustment score on the California Test of Personality, (*b*) the Taylor Manifest Anxiety Score, and (*c*) the self-confidence scale score of the Bernreuter Personality Inventory. Only the self-confidence score showed a significant difference between the two groups, with Negroes scoring in the direction of lack of self-confidence.

Correlations between the two measures of intelligence and the three measures of personality were all significant for the Negro group. The highest relationships for the Negro group were obtained between pattern analysis and lack of self-confidence (−.88), between the combined intelligence score and lack of self-confidence (−.79), and between pattern analysis and total adjustment (.73).

For the white group, significant correlations were obtained between pattern analysis and lack of self-confidence (−.46), between combined intelligence score and lack of confidence (−.50), and between combined intelligence score and total adjustment (.35).

The differences between the correlations for the Negro and white groups on the same correlated variables showed significance beyond the .01 level between correlations with respect to total adjustment and pattern analysis, lack of self-confidence and the combined intelligence score, and lack of self-confidence and pattern analysis. The difference between correlations for anxiety and the combined intelligence score was significant at the .05 level. In each case the correlation associated with the Negro group was larger.

DISCUSSION AND CONCLUSIONS

The personality tests selected were chosen on face value as possibly associated with ability to score on intelligence tests. Of the scales available, psychological adjustment, manifest anxiety, and self-confidence seemed to fit this criterion. High correlations between the measures of intelligence and the personality variables were obtained. That these measures of personality in Negroes were more associated with ability to score on intelligence tests than they were with whites was demonstrated by the significant differences between the correlations for the two groups.

Part of the relationship between the personality measures and intelligence might be explained in terms of the wide spread of intelligence in the sample (intelligence measures ranged from 70 to 139). Taylor (1955) cited evidence of this relationship. However, the range and standard deviations of the measures used are not so different for the

Negro and the white groups as to compel the conclusion that the results can be fully accounted for in this way.

The hypotheses tested were confirmed: the Negro group scored lower on the ACB than did the whites; the Negro group manifested a greater lack of self-confidence than did the whites; and all three measures of personality showed higher correlations with ACB scores for the Negro group than they did for the white group.

Although statements of causality cannot be made from the data presented, the evidence can be seen as warranting further research on the proposition that Negroes as a group, lacking support from pride in significant historical achievement, and developing in an environment of negative experiences, incorporate intellectually defeating personality traits that play a significant role in their ability to score on measures of intelligence.

REFERENCES

ANASTASI, ANNE, & D'ANGELO, RITA Y. A comparison of Negro and white preschool children in language development and Goodenough Draw-a-Man IQ. *J. genet. Psychol.*, 1952, **81**, 147–165.

BERNREUTER, R. G. *Manual for the personality inventory.* Stanford: Stanford Univer. Press, 1935.

BOYD, G. F. The levels of aspiration of white and Negro children in a nonsegregated elementary school. *J. soc. Psychol.*, 1952, **36**, 191–196.

GILLILAND, A. R. Socio-economic status and race as factors in infant intelligence test scores. *Child Develpm.*, 1951, **22**, 271–273.

KARON, B. P. *The Negro personality.* New York: Springer, 1957.

KLINEBERG, O., ET AL. On race and intelligence: A joint statement. *Amer. J. Orthopsychiat.*, 1957, **27**, 420–422.

LONG, H. H. The relative learning capacity of Negroes and whites. *J. Negro Educ.*, 1957, **26**, 121–134.

MCGURK, F. C. J. "Psychological tests": A scientist's report on race differences. *U.S. News World Rep.*, 1956, Sept. 21, 92–96.

MONTAGUE, E. K., WILLIAMS, H. L., LUBIN, A., & GIESEKING, C. F. The use of army tests for the assessment of intellectual deficit. *USA res. Rep.*, 1957, WRAIR–52–57.

NORTH, R. D. The intelligence of American Negroes. *Res. Rep. Anti-Defamation League B'nai B'rith*, 1956, **3**, 2–8.

SHUEY, AUDREY, M. *The testing of Negro intelligence.* Lynchburg, Va.: Bell, 1958.

SONTAG, L. W., BAKER, C. T., & NELSON, VIRGINIA. Personality as a determinant of performance. *Amer. J. Orthopsychiat.*, 1955, **25**, 555–562.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc., Psychol.*, 1953, **48**, 285–295.

TAYLOR, JANET A. The Taylor Manifest Anxiety Scale and intelligence. *J. abnorm. soc. Psychol.*, 1955, **51**, 347.

THORPE, L. P., CLARK, W. W., & TIEGS, E. W. *California Test of Personality.* Los Angeles: California Test Bureau, 1953.

TRUMBULL, R. A study in relationships between factors of personality and intelligence. *J. soc. Psychol.*, 1953, **38**, 161–173.

(Early publication received December 2, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 151–154

PATTERNS OF PERMISSIVENESS AMONG PRELITERATE PEOPLES

E. TERRY PROTHRO
American University of Beirut

Is PERMISSIVENESS in child rearing a general characteristic of maternal behavior—a dimension, a trait? Or does permissiveness of mothers vary with the type of child behavior? Studies of American mothers do not seem to yield consistent answers on this point. Sewell (1955) interviewed a sample of middle-class rural Wisconsin mothers and found little relationship among various kinds of permissiveness. Sears (1957) interviewed working class and middle-class mothers in two New England suburbs, and found a general permissiveness factor. Such inconsistency of results in studies of specific communities seems fairly typical of recent American surveys of child rearing practices. Reports on practices in Illinois (Davis, 1952), Massachusetts (Maccoby & Gibbs; 1954), Oregon (Littman, Moore, & Pierce-Jones, 1957), and California (White, 1957) differ to such a degree that most students of maternal behavior would concede with Havighurst and Davis (1955, p. 441) that "we should not attempt to generalize concerning child-rearing . . . from a sample in one part of the country, even if it is a representative sample."

As Whiting has pointed out (Whiting, 1953), contradictory results in studies of individual communities often may be resolved by utilizing cross-cultural data. Such data reduce the possibility of negative findings attributable to restricted range of behavior, or of positive findings attributable to temporary or local factors. These considerations highlight the importance of the study by Whiting and Child (1953) of the child rearing norms of cultures from all parts of the world. From this study they concluded (p. 117) that among nonliterate peoples "the [child rearing] practices of a society for one system of behavior are almost entirely independent of its practices with respect to other systems of behavior." Cultural norms of child rearing, then, do not differ with respect to permissiveness in general, but differ with respect to permissiveness for sex behavior, permissiveness for dependent behavior, etc. This assertion is, as they pointed out, one of considerable theoretical and practical importance. It is with that assertion, and the data on which it is based, that the present study is concerned.

Source of Data

In their cross-cultural study, Whiting and Child (1953) gathered data on child rearing in 75 cultures by utilizing the Human Relations Area Files and some other "pertinent articles or monographs which had not yet been entered in the file." This information was placed in the hands of experienced judges, who assigned ratings (whenever possible) to each culture with respect to its indulgence or permissiveness for five "systems of behavior"—oral, anal, sexual, dependence, and aggression. For each system, separate judgments were made of initial satisfaction, age at serious socialization, and amount of socialization anxiety. Finally, each culture was rated (whenever possible) with respect to certain customs relating to illness. All of these ratings are reported in the appendix of *Child Training and Personality* (Whiting & Child, 1953, pp. 338–346).

Although they were interested in the interrelations of the child rearing practices, Whiting and Child did not carry out a factor analysis of their data. For the purposes of this study, it was decided to re-examine their data by the use of factor analysis. Two analyses were carried out. Thurstone's centroid method was used, and axes were rotated graphically. The first analysis was based on ratings assigned those 22 cultures on which fairly complete information was available. The following cultures were included because they had all been rated on 30 of the practices: Chagga, Comanche, Dahomeans, Hopi, Kwakiutl, Kwoma, Lepcha, Lesu, Manus, Marquesans, Navaho, Ontong-Javanese, Papago, Pukapukans, Siriono, Tanala, Tenino, Teton, Western Apache, and Wogeo. After the initial analysis had been completed, and it had been established that there were several dimensions of child rearing, a factor analysis was made of data from 36 cultures to determine whether the same dimensions would emerge with a larger sample. In the second analysis, the 22 cultures of the first analysis were included, and added to them were 14 cultures on which ratings were available for 12 child rearing practices and 7 illness practices. By reducing the number of practices included in the analysis it was possible to include a larger number of cultures. The additional cultures were: Ainu, Arapesh, Azande, Bena, Chamorro, Chenchu, Chiricahua, Kurtatchi, Kutenai, Maori, Samoans, Thonga, Trobrianders, and Yakut.

Results

Rotated factor loadings for each of the analyses are shown in Table 1. It can be seen that the three factors emerging from the first analysis correspond quite closely to those of the second analysis.

The factor identified as *F2′* in each of the

TABLE 1
ROTATED FACTOR LOADINGS

Variable	First Analysis (22 Cultures)			Second Analysis (36 Cultures)		
	F1′	*F2′*	*F3′*	*F1′*	*F2′*	*F3′*
Initial Indulgence						
A Oral	−11	−65	−08	−08	−63	−28
B Anal	01	25	−78	−22	32	−71
C Sexual	00	61	−21	−42	40	−20
D Dependence	−07	−58	11	13	−49	00
E Aggression	−56	17	−22	−52	−05	−24
Socialization Anxiety						
G Oral	44	43	−16	27	54	00
H Anal	02	−11	75	07	−04	85
I Sexual	07	−73	12	51	−53	17
J Dependence	33	24	−50	17	42	−40
K Aggression	48	−15	16	66	09	14
Age at Socialization						
M Weaning	11	−55	10	−05	−46	−27
N Toilet Training	01	−22	−63	—	—	—
Q Independence	−18	−14	54	20	−47	51
Illness Explanations						
a Oral	49	48	−13	08	52	−21
b Anal	68	−02	29	54	02	07
c Sexual	44	−10	−22	42	−06	−11
d Dependence	03	73	01	−04	57	06
e Aggression	71	01	−34	52	32	−26
Performance Therapy						
f Oral	39	−09	18	—	—	—
g Anal	55	−04	14	—	—	—
i Dependence	37	−12	06	—	—	—
j Aggression	41	−06	15	—	—	—
Avoidance Therapy						
k Oral	93	−10	09	—	—	—
l Anal	71	−04	07	—	—	—
m Sexual	36	−29	−15	—	—	—
n Dependence	12	02	−45	—	—	—
o Aggression	30	−07	15	—	—	—
Fear of						
p Human beings	38	−21	16	38	−07	24
q Spirits	−34	34	13	−21	23	18

analyses could be called an *orality–sexuality* factor. The culture high on this factor characteristically indulges the infant's oral needs, weans late, evokes little anxiety over weaning, indulges the infant's expressions of dependency, but also is nonpermissive of infantile sexuality and produces considerable anxiety over sex training. Cultures at the other extreme are permissive regarding sex but nonindulgent of the infant's oral and dependency needs. They are also characterized by "oral explanations of illness" and "dependency explanations of illness." That is, these cultures typically explain illness as caused by ingesting something, by verbal spells (oral) or by the soul leaving the body, or as an evil spirit entering the body (dependency).

The factor called *F3′* in both analyses could be named an *independence–anality factor*. The culture high on this factor begins toilet training early and evokes much anxiety over the training, but it begins serious independence training comparatively late. Cultures at the other extreme are indulgent about the infant's anal activity and late in toilet training, but they begin serious independence training early and produce much anxiety in the process.

The third factor, *F1′* in these analyses, might be called an *aggression–hypochondria factor*. Almost all of the practices relating to illness had loadings on this factor, so it was called "hypochondria" to emphasize the generalized concern with illness. The only child rearing practices with substantial loadings on this factor in both analyses were the two pertaining to permissiveness for aggression. Cultures high on this factor permitted little aggression in small children, produced anxiety in the process of socializing aggression, and were generally high in the use of various therapies and explanations for illness. Treatment of infantile sexuality had a significant loading on this factor in the second analysis, but no loading at all in the first analysis.

DISCUSSION

The child rearing norms of these cultures cannot be described as permissive or nonpermissive. Rather, they must be described as more or less permissive on orality-sexuality, on anality-independence, and on aggression. It is not even accurate to refer to the "permissiveness for dependency" in a given culture, for dependence is not unifactorial. Permissiveness for dependency behavior in the nursing infant is not correlated with permissiveness for dependency at the time of serious socialization, or the degree of anxiety produced by such independence training. This dual nature of dependency was noted by Whiting and Child (1953, p. 109).

While it cannot be asserted that permissiveness is wholly generalized, neither does it seem correct to say that permissiveness is specific to the system of behavior–oral, sexual, anal, dependence, or aggression. There are among these systems definite clusters or patterns in the cultural norms (cf. Littman & Moore, 1957, p. 703). The assertion of Whiting and Child that "there is no consistent relation between the way one system of behavior is socialized and the way any other system of behavior is socialized" (Whiting & Child, 1953, pp. 115–116) overemphasizes the specificity of norms. They computed intercorrelations among

ratings on the five systems for initial indulgence, and repeated this for severity of socialization and age of socialization. Because ratings for many cultures were incomplete, the maximum number of cultures used in computing these correlations was 36. No correlations were computed between ratings of one system on initial indulgence and ratings of other systems on severity of socialization or age at socialization. The "no consistent relation" conclusion was based on the observation that significant correlations among those computed were few, and that positive and negative correlations were about equally prevalent. Such observations do not seem to justify their conclusion. Patterns in the data could have been found by utilizing other techniques.

The bipolarity of the *oral-sexual* factor could be interpreted in any one of a number of ways. The psychoanalytically oriented might argue that the children, and by extrapolation the culture, are fixated at either the oral or the phallic stage of psychosexual development. Others might approach the problem from the standpoint of the adult and say that the mother finds nursing and sex to be alternative types of gratification. In a culture where sex activity is encouraged and accepted, the mother may find prolonged nursing a sex handicap. Thus Linton (1939, pp. 164–165) describes Marquesan mothers who disliked nursing because of a fear that their breasts, considered important to sex play in that culture, might be "spoiled." The converse of this situation seems to exist among some Arab mothers, for whom sex is shameful and frightening, but who confess blushingly that they find nursing their child to be pleasurable (Ammar, 1954, p. 190; Wakim, 1956, p. 50).

The opposition between toilet training and independence training in the *independence-anal* factor is not difficult to understand when it is recognized that much of "independence training" consists of leaving a child to his own resources (Whitney & Child, 1953, p. 97). If the two-and-a-half year old child is allowed considerable freedom in developing his defecation habits, he may well be required at three-and-a-half to develop his own habits of play and self-protection. There is also a possibility that *ecological variables* play a part in opposing "anal" and "independence" indulgence. In a Lebanese mountain-village surrounded on three sides by canyons, and covered in winter with deep snow, a mother may be compelled to restrict the outdoor activity of her three year old. At the same time, the confined quarters would inhibit a casual attitude toward toilet training. The environment would compel parents to delay or ignore independence training and to inhibit "anal indulgence." Beduoins, on the other hand, live more in open country, and are more mobile, so that sanitation is less of a problem. At the same time, it is more feasible for the mother to let the child shift for himself outside the family tent. Their environment would encourage late toilet training and early independence training. Hence, the physical environment may in some instances contribute to the opposition between toilet training and independence training.

Although the *aggression-hypochondria* factor seems to be the same in both analyses, there are two reasons for caution in asserting the presence of such a factor. First, therapy ratings were not included in the second analysis. More important, however, is the very real possibility that the interrelations between the various illness practices may have been an artifact of the scoring procedure. The judges who rated the cultures were instructed to give ratings of plus when the ethnographic material on a culture mentioned a certain practice and minus when there was mention of the absence of a practice *or no mention of the practice at all*. If one ethnographer, then, noted carefully a wide variety of beliefs and practices regarding illness, and another noted only the most striking ones the second culture would be scored as lacking practices found in the first. Whiting (1954 p. 529) has noted that the procedure may yield spurious correlations, and there is a definite possibility that such spurious correlations could have affected the results of our analysis. The "hypochondria" factor might represent ethnographer's generalized concern and lack of concern with illness rather than the culture's concern or lack of it.

SUMMARY

In order to ascertain the generality of norms of permissiveness in child rearing, data published by Whiting and Child (1953) on primitive cultures from all parts of the world were examined by the use of factor analysis. It was found that there were three relatively unrelated dimensions of permissiveness: orality-sexuality, independence-anality, and aggression. Sexual permissiveness was associated with nonpermissiveness on oral matters, and oral permissiveness with nonpermissiveness on sex. Similarly, permissiveness in independence training was negatively related to permissiveness in toilet training. There was some evidence for a general hypochondria factor as a part of the aggression factor. It was pointed out that the patterning of child rearing norms might be explained in terms of conscious motives of parents, ecological variables, or even the concerns of individual ethnographers.

REFERENCES

AMMAR, H. *Growing up in an Egyptian village*. London: Routledge & Kegan Paul, 1954.

DAVIS, A., & HAVIGHURST, R. J. Social class and color differences in child rearing. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley, (Eds.), *Readings in social psychology*. New York: Holt, 1952. Pp. 539–550.

HAVIGHURST, R. J., & DAVIS, A. A comparison of the Chicago and Harvard studies of social class differences in child rearing. *Amer. sociol. Rev.*, 1955, **20**, 438–442.

LINTON, R. Marquesan culture. In A. Kardiner (Ed.), *The individual and his society*. New York: Columbia Univer. Press, 1939. Pp. 137–196.

LITTMAN, R. A., MOORE, R. C. A., & PIERCE-JONES, J. Social class differences in child rearing: A third community for comparison with Chicago and Newton. *Amer. sociol. Rev.*, 1957, **22**, 694–704.

MACCOBY, ELEANOR E., GIBBS, PATRICIA K., ET AL. Methods of child rearing in two social classes. In W. E. Martin & C. B. Stendler (Eds.), *Readings in child development*. New York: Harcourt, 1954. Pp. 380–396.

SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child rearing*. Evanston: Row, Peterson, 1957.

SEWELL, W. H., MUSSEN, P. H., & HARRIS, C. W. Relationships among child training practices. *Amer. sociol. Rev.*, 1955, **20**, 137–148.

WAKIM, SUAD G. Child care in Mieh-Mieh. Unpublished master's thesis, American Univer. Beirut, Lebanon, 1956.

WHITE, MARTHA S. Social class, child rearing practices, and child behavior. *Amer. sociol. Rev.*, 1957, **22**, 704–712.

WHITING, J. W. M. The cross-cultural method. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge: Addison-Wesley, 1954. Pp. 523–531.

WHITING, J. W. M., & CHILD, I. L. *Child training and personality*. New Haven: Yale Univer. Press, 1953.

(Received February 5, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 1, 155–158

CHARACTERISTICS OF UNTESTABLE CHRONIC SCHIZOPHRENICS[1]

HAROLD WILENSKY AND LEONARD SOLOMON[2]
Franklin D. Roosevelt Veterans Administration Hospital, Montrose, New York

RESEARCH studies in schizophrenia which employ clinical tests or laboratory behavior as dependent variables invariably report that some proportion of the *Ss* are untestable or uncooperative. It is tacitly assumed that conclusions drawn from such studies will be delimited to the population of testable patients. Unfortunately, however, the results of such studies generate clinical theory which tends, by and large, to neglect the limitations imposed by the particular character of the testable sample.

It seems desirable to establish criteria for defining "untestability" and to determine which factors are likely to affect this attribute. Moreover, the measure of testability may be a variable worthy of study in its own right in that there is evidence to indicate that it has significant behavioral correlates. Mednick and Lindsley (1958) reported that the rate of response under operant conditioning was related to testability and could have implications for the patient's responsiveness to the environment.

Zubin, Windle, and Hamwi (1953) reported that 25% of the chronic schizophrenics used in the Columbia-Greystone project were untestable on more than one third of a test battery including cognitive (IQ), perceptual, and motor components. Unfortunately, the untestability rates for specific test components were not indicated. King (1954), using relatively simple motor tasks, reported the rate of untestability to be as low as 10% for his chronic schizophrenic sample (median length of hospitalization of 9.1 years). Interestingly enough, Mednick and Lindsley (1958), with an extremely chronic sample (median length of hospitalization of 16 years) on a battery of Wechsler-Bellevue and Rorschach, reported that 55% were untestable. It would appear that the proportion of untestable patients obtained is determined by the characteristics of the patients, task complexity, and social situational factors.

The present study was essentially exploratory in nature and sought to examine the factors producing untestability, to estimate the stability of such behavior, and to compare the differences between testable and untestable patients in order to make explicit the nature of the limitations to be placed upon theoretical generalizations.

PROCEDURE

The sample consisted of the entire population of two wards of 101 chronic male schizophrenic patients. The patients ranged in age from 26 to 70 years with periods of continuous hospitalization ranging from one to 29 years.

Attempts were made to administer individually a Rorschach and an abbreviated WAIS[3] to each patient on three separate occasions, one month apart. The patients were randomly assigned to an experienced clinical psychologist or to relatively inexperienced psychology trainees for all three testing occasions. The tests were administered in a private office adjacent to the ward dayroom. If a patient refused to come to the office or refused to respond to the test questions, he was encouraged to participate by suggesting that the examiner was interested in his progress and that all patients on his ward were being examined.

A number of demographic indices were derived from case history material: age, race, place of birth, education, and length of hospitalization. Indices of current performance were obtained from a variety of sources and by different techniques: diagnosis, mental health ratings, rates of social interaction, and tapping test performance. They are listed and operationally defined below:

Diagnosis. This classification was derived from a mental status examination administered by two psychiatric residents.[4] The three diagnostic categories employed were hebephrenic, paranoid, and other (catatonic, mixed, or simple).

Mental health ratings. Three social psychology trainees, functioning as ward observers, rated each patient on a 6-point scale which focused upon the patient's ability to assume independent responsibility in caring for himself. The scale ranged from ability

[1] Part of this paper was read at the Eastern Psychological Association meetings at Atlantic City, New Jersey, April, 1959.

[2] The authors would like to acknowledge the help given by Alfred Drelich in the statistical analysis of the data. We are also indebted to the Social Psychology Unit for the use of its social interaction data and health ratings, and to the psychology trainees who participated in the repeated testing of patients.

[3] Five subtests were selected, representing each of five factors extracted from a factor analysis of the WAIS by Cohen (1957). They were Information, Digit Span, Picture Completion, Block Design, and Digit Symbol.

[4] Joseph Dorsey and Ralph Baker made the psychiatric diagnoses based upon mental status examinations. Each psychiatrist independently examined approximately half of the cases.

to live outside the hospital and work (score one) to requiring maximum supervision for his daily activities in the institution (score six). These ratings were made solely on the basis of observed patient behavior on the ward. The interobserver reliability coefficients for the three possible pairs were .73, .77, and .82.

Rates of social interaction. The social psychologists, acting as ward observers, recorded the frequency of patient to patient interactions on the ward over a 3-month period. There were 55 one-hour periods of observation made at random times throughout the day. A measure of the interaction rate for each patient was based upon the following ratio: number of patients' social contacts/number of observation periods in which patient was present.[5]

Tapping test. About 6 months after the administration of the WAIS and Rorschach, the patients received a simple motor test of finger tapping. Conditions differed, however, in that the tapping test was given on the ward, whereas the clinical tests were administered in a private office.

RESULTS

The first testing occasion (Rorschach and WAIS) was used to define testability. Fourteen patients refused to enter the examiner's office despite the encouragement by the psychologist. These patients made clear their intentions not to take tests, and were labeled *untestable refused*. Another 13 patients entered the examiner's office but could not be tested. These patients did not refuse explicitly. Most of them simply replied "I don't know," or made no response to questions. Some were incoherent or irrelevant in their responses. Because they appeared incapable of dealing with the tasks, these patients were labeled as *untestable confused*. The rationale for this classification is based on the assumption of a motivational difference between the confused and refused untestable patients. The confused untestables seemed willing but unable to produce scorable records, whereas the untestable refused group appeared unwilling to perform. Obviously, no judgment could be made as to the ability of the latter group to respond to tests.

Of the remaining patients, 64 gave adequate WAIS and Rorschach protocols. Ten produced only one scorable test (either WAIS or Rorschach). These patients were classified as *testable*. In the subsequent data analyses, the testable, untestable confused, and untestable refused groups were compared with regard to situational, demographic, and current performance variables.

[5] In order to test the interobserver reliability of the interaction rate, two observers made independent judgments with respect to the type and role of a patient's participation for 65 interactions. Using an index of reliability for nominal categories known as Kappa (Cohen, 1960) which represents the proportion of observer agreement after chance agreement is removed, we obtained $K = .71$, significant at. <01 level.

TABLE 1

CHARACTERISTICS AND NUMBER OF PATIENTS IN THE TESTABILITY TYPES

Variable	Testable	Untestable Confused	Untestable Refused	*P* Value
Examiner experience:				
Trainee	36	7	7	ns
Staff	38	6	7	
Birthplace:				
Native	69	7	12	<.05[a]
Foreign born	5	6	2	
Diagnosis:				
Paranoid	25	1	10	<.005[b]
Hebephrenic	37	6	3	
Other	12	6	1	
Total	74	13	14	
Central Tendencies:				
Median age	39.5	42.0	46.5	<.02[c]
Median age 1st hospitalization	26.2	26.0	31.0	<.10[c]
Median years continuous hospitalization	9.8	14.5	10.5	<.05[c]
Median education grade	10.0	8.5	8.0	<.10[c]
Median social interaction rate	.063	.033	.050	<.10[c]
Mean mental health rating	3.75	4.88	4.13	[d]

[a] Fisher's exact test.

[b] Although the requirements of expected cell frequencies of 5 for using chi square were not met, the p value obtained suggests that the null hypothesis may be rejected.

[c] The Kruskal-Wallis test of differences in rank order was used for comparisons among groups.

[d] The untestable confused group significantly differed in mean and in variance from the testable group ($t = 5.12$; $F = 2.90$) and from the untestable refused group ($t = 2.19$; $F = 2.67$). No significant differences were found between the untestable refused and the testable group ($t = 1.24$; $F = 1.08$).

Demographic Variables

Race. No differences in testability related to race were found. Of the fifteen Negroes in the total sample, one refused and one was classified as confused.

Place of birth. Foreign born patients appeared with greater frequency in the untestable confused group. It was likely that language difficulty contributed to the apparent confusion in this group of untestable patients. (See Table 1.)

Age and hospitalization. Age, age at onset of illness, length of continuous hospitalization, and education yielded skewed distributions which were analyzed by means of the Kruskal-Wallis test of differences in rank order (Senders, 1958). With regard to age, the untestable refused group emerged as the oldest. Moreover, the data suggest that their onset of illness may have taken place

at a later age. The median period of continuous hospitalization for the refused cases was relatively low. By contrast, the untestable confused group was closer to the median of the testable sample in both chronological age and age of onset of illness. Their period of continuous hospitalization, however, was the longest. In fact, the majority of patients in this confused group have never been separated from a hospital since their first admission. Had the samples been larger, both untestable groups might have shown less education than the testable group. (See Table 1 for medians and p values.)

Current Performance and Behavior

Diagnosis. Inspection of the distribution of diagnoses revealed that the untestable refused group contained a greater number of paranoid schizophrenics. The untestable confused group contained fewer paranoid patients. (See Table 1 for diagnostic distributions.)

Mental health ratings

Bartlett's test of homogeneity of variance in mental health ratings among the three groups yielded a significant chi square (38.91). Since the assumptions underlying analysis of variance could not be met, the three groups were compared with respect to mental health ratings by means of t tests. The poorest ratings occurred in the untestable confused group, which also was less variable than either of the other two samples. No significant differences were found between the untestable refused and the testable group in either mean tendency or variability. The variances for the untestable confused, untestable refused, and the testable groups were .39, 1.06, and 1.15 respectively. (See Table 1 for mean health ratings.)

Rates of social interaction. With respect to social interaction, the findings, although not significant, suggested trends parallel to those observed in the mental health ratings. The untestable confused group seemed to manifest somewhat lower rates of interaction as well as a restricted range when compared with the other two groups. The entire confused group was a bit below the interaction rate median for the entire group. (See Table 1 for median rates.)

Miscellaneous Variables

Examiner experience. The effects of examiner experience did not result in any significant differences in testability. The numbers of patients in each of the testability categories for the staff and trainee psychologists are presented in Table 1.

Consistency. The consistency of testability was determined by repeating the WAIS and Rorschach examinations on two more occasions. Among initially untestable patients, approximately 90% continued to be untestable on clinical tests. Of the initially testable sample, 9 patients refused one or more later examinations. Thus, there was a high degree of consistency in testability. Eighty-eight per cent of the patients were consistent in their accepting or rejecting tests on all three occasions.

Tapping test. The tapping test performance was not included as a criterion of testability because of the different time of administration and examination conditions. It was of interest to note that the percentage of untestable patients dropped to 9% with this simple motor task, which was administered in full view of other patients on their ward. Approximately two-thirds of the untestable patients accepted and performed the tapping test.

DISCUSSION

The results suggest that untestable patients do not compose a homogeneous category. Untestable patients can meaningfully be divided into at least two major subgroups according to *level of ability* in test performance and *motivation* to accept the task. Of the total untestable sample, 13 were cooperative but apparently confused, and 14 explicitly refused to be tested.

Significant differences among the groups were demonstrated for a number of variables such as diagnosis, age, years of continuous hospitalization, and mental health ratings. Although social interaction and age at onset did not attain significance, the pattern of differences was similar to the other variabiles which essentially represented aspects of a more general concept of mental health.

The untestable confused group appeared to be more disturbed on the basis of a variety of criteria: (*a*) their long length of continuous hospitalization, (*b*) their low health ratings, and (*c*) possibly their degree of social withdrawal. Even with the simple finger tapping task, one-third of this group was unable to respond adequately. The characterization of untestable patients made by Zubin et al. (1953) that they are "essentially deteriorated" with poor prognoses would appear to be specifically applicable only to the confused category.

In contrast to this, those patients who explicitly refused to submit to testing were indistinguishable from the testable patients in terms of degree of health, length of continuous hospitalization, or rates of social activity. The untestable refused group differed from the over-all sample in that they were the oldest group, and they were likely to be diagnosed as paranoid. Thus, although this category of patients was untestable, they did not reflect the descriptive image offered by either Zubin et al. (1953) or Mednick and Lindsley

(1958). They could not be considered any more deteriorated or unresponsive to their environment than the testable sample.

Twelve per cent of the sample exhibited fluctuations in type of test performance when the first testing was compared to either the second or third occasion. In effect, the total sample of untestable patients increases by 9% if the entire three testing occasions are used as a criterion. For those investigators who employ a repeated measures design, there are further limitations imposed by this increase in the untestable sample.

On the basis of incidental data, it may be suspected that for different tasks or for different test conditions, testability rates may be substantially altered. For example, the use of simple operant tasks or public testing situations may have a salutary effect on the testability rate. Paranoid patients in particular may be more likely to refuse to enter a private office than to accept tests under visibly safe public conditions.

It is clear from the data that the testable sample, on any test occasion, represents a selectively biased sample of the entire male VA chronic schizophrenic population.

SUMMARY

Attempts were made to examine the entire population of two wards of chronic schizophrenic patients with an individually administered Rorschach and an abbreviated WAIS. A finger tapping test was administered at a later date on the ward. Patients were classified into three groups according to their willingness to accept testing and ability to perform. Twenty-seven per cent of the patients were untestable. The untestable chronic male schizophrenic patients seemed to represent a population qualitatively different from the testable population.

The present investigation therefore affirms the necessity for widening research efforts to include a more articulate description of the "silent sample," untestable schizophrenics.

REFERENCES

COHEN, J. The factorial structure of the WAIS between early adulthood and old age. *J. consult. Psychol.*, 1957, **21**, 283–290.

COHEN, J. A coefficient of agreement for nominal scales. *Educ. psychol. Measmt.*, 1960, **20**, 37–46.

KING, H. E. *Psychomotor aspects of mental disease*. Cambridge, Mass.: Harvard Univer. Press, 1954.

MEDNICK, MARTHA T., & LINDSLEY, O. W. Some clinical correlates of operant behavior. *J. abnorm. soc. Psychol.*, 1958, **57**, 13–16.

SENDERS, VIRGINIA L. *Measurement and statistics*. New York: Oxford Univer. Press, 1958.

ZUBIN, J., WINDLE, C., & HAMWI, VIOLET. Retrospective evaluation of psychological tests as prognostic instruments in mental disorders. *J. Pers.*, 1953, **21**, 342–355.

(Received April 30, 1959)

Neurology and Psychiatry

Section VIII of Excerpta Medica

The Neurology Section includes abstracts of all articles of neurological interest in addition to general clinical neurology and experimental neurophysiology.

The section on Psychiatry covers the topics usually considered within the scope of the word. In addition the following special branches are dealt with fully: psychoanalysis and individual and analytical psychology; neurology, criminology, sexuality, and drug addiction as related to psychiatry; psychosomatic medicine; social and industrial psychology and psychiatry; vocational guidance, and mental testing; psychometrics; personality testing and the Rorschach test; heredity and statistical studies as they apply to psychiatry. The subjects of mental defect and epilepsy are fully reported.

Annual volume of 1,000 pp., including an index classified by author and subject.

Subscription price $42.50 per annual volume

EXCERPTA MEDICA FOUNDATION

The New York Academy of Medicine Building

2 East 103 Street New York 29

PSYCHOLOGICAL MONOGRAPHS

No. 480. Similarities in the Structure of Extreme Social Attitudes
by Irvine L. Taylor
Price $1.50

No. 483. Relative Effectiveness of Psychotherapeutic Programs: A Multicriteria Comparison of Four Programs for Three Different Patient Groups
by G. W. Fairweather, R. Simon, M. E. Gebhard, E. Weingarten, J. L. Holland, R. Sanders, G. B. Stone, and J. E. Reahl
Price $1.00

Order from:
AMERICAN PSYCHOLOGICAL ASSOCIATION
Dept. JASP, 1333 16th Street, N.W.
Washington 6, D. C.